

STIC-Biotech/ChemLib

79425

09/782,874

From: Chan, Christina
Sent: Tuesday, November 05, 2002 9:32 AM
To: Helmer, Georgia; STIC-Biotech/ChemLib
Subject: RE: Please Rush Sequence Search-09/782,874--

Point of Contact:
Mona Smith
Technical Information Specialist
CM1 6A01
Tel: 308-3278

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

CRFE

-----Original Message-----

From: Helmer, Georgia
Sent: Monday, November 04, 2002 6:24 PM
To: Chan, Christina
Subject: Please Rush Sequence Search-09/782,874--

Could you please do a sequence search and an oligo search on SEQ ID No 1 and 2?

On the results for the oligo searches, could you please print out more of the results? Like up to the first 50?

This is for the inhouse and commerce data bases. (no interference search though)

I need this case for this biweek.

Thanks in advance, Georgia L. Helmer Ph.D.
Patent Examiner
Crystal Mall 1, 9D14
AU 1638
703-308-7023
Georgia.Helmer@USPTO.gov

16 coning bases
= 0.42%
0.42%

510 1 = 37.31 nuc

Searcher: M. Smith
Phone: _____
Location: _____
Date Picked Up: 11/04/02
Date Completed: 11/06/02
Searcher Prep/Review: 10
Clerical: _____
Online time: 15

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: 2
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 19:16:42 : Search time 142 seconds
(without alignments)
8057.824 Million cell updates/sec

Title: US-09-782-874-1
Perfect score: 3731
Sequence: 1 GAATATCTTCTTACTTACTT.....AGTTTCATCTTCTTCTTAAA 3731

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTOUT_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3731	100.0	3731	4	US-08-811-583-1
2	57.2	1.5	906	1	US-08-100-874-1
3	53.4	1.4	7218	1	US-08-232-463-14
4	40	1.1	162450	4	US-09-345-882-1
5	36.8	1.0	2277	1	US-08-676-967-2
6	36.8	1.0	2277	1	US-08-676-974-2
7	36.8	1.0	2277	2	US-09-058-487-2
8	36.6	1.0	7218	1	US-08-232-463-14
9	36.4	1.0	3191	4	US-08-936-165A-194
10	35.4	0.9	17341	4	US-09-415-946-1
11	35	0.9	1248	4	US-09-134-001C-342
12	35	0.9	1275	2	US-08-911-445-4
13	35	0.9	1275	4	US-09-182-983-4
14	35	0.9	1458	2	US-08-911-445-19
15	35	0.9	1458	4	US-09-182-983-19
16	35	0.9	1796	2	US-08-911-445-17
17	35	0.9	1796	4	US-09-182-983-17
18	35	0.9	1845	2	US-08-911-445-5
19	35	0.9	1845	4	US-09-182-983-5
20	35	0.9	2046	2	US-08-911-445-6
21	35	0.9	2046	4	US-09-182-983-6
22	35	0.9	5613	2	US-08-463-418-1
23	35	0.9	111282	4	US-09-754-250-3
24	34.8	0.9	1028	4	US-08-858-207A-222
25	34.8	0.9	6002	4	US-09-345-882-4
26	34.8	0.9	19446	4	US-08-961-527-51
27	34.6	0.9	1883	4	US-08-991-677-3

28	34.6	0.9	19390	4	US-08-961-527-86	Sequence 86, Appl
29	34.4	0.9	738	4	US-09-004-838-62	Sequence 62, Appl
30	34.2	0.9	1003	4	US-09-221-017B-395	Sequence 395, Appl
31	34.2	0.9	2694	1	US-08-465-995A-1	Sequence 1, Appl
32	34.2	0.9	2694	2	US-08-465-994C-1	Sequence 1, Appl
33	34.2	0.9	2694	2	US-08-966-145-1	Sequence 1, Appl
34	34.2	0.9	2760	1	US-08-101-593-1	Sequence 1, Appl
35	33.8	0.9	258	4	US-09-345-882-21	Sequence 21, Appl
36	33.8	0.9	1867	4	US-09-377-557-3	Sequence 3, Appl
37	33.8	0.9	2889	4	US-09-134-001C-1660	Sequence 1660, Ap
38	33.6	0.9	2870	4	US-09-134-078-59	Sequence 59, Appl
39	33.6	0.9	3707	1	US-08-118-101A-1	Sequence 1, Appl
40	33.6	0.9	10223	4	US-08-961-527-73	Sequence 73, Appl
41	33.4	0.9	1939	1	US-08-212-188-3	Sequence 3, Appl
42	33.4	0.9	1939	3	US-08-970-725-3	Sequence 3, Appl
43	33.4	0.9	1939	5	PCT-US95-02708-3	Sequence 3, Appl
44	33.4	0.9	2026	3	US-08-755-587-26	Sequence 26, Appl
45	33.2	0.9	3636	4	US-09-134-001C-1983	Sequence 1983, Ap

ALIGNMENTS

RESULT 1

US-08-811-583-1
; Sequence 1, Application US/08811583
; Patent No. 6218142

GENERAL INFORMATION:

APPLICANT: Wasseneqger, Michael
APPLICANT: Riedel, Leonhard
APPLICANT: Schiebel, Winfried
APPLICANT: Sanger, Heinz
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
RNA-DIRECTED RNA POLYMERASE (RDRP)
TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/811,583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090

INFORMATION FOR SEQ ID NO:

1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3731 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Tomato
FEATURE:
NAME/KEY: CDS
LOCATION: 194..3535

US-08-811-583-1

Db 2101 TGATCTTTTGAAGGAGGAGGCTTTGAATTCATGCTCTCTGGAGAGACACTAATAT 2160
QY 2161 TCTCAAGGCAATGCTAACTGTGTTTAAAGCCTGATGCTGAGCCCTTTCTTTTCAATGAT 2220
Db 2161 TCTCAAGGCAATGCTAACTGTGTTTAAAGCCTGATGCTGAGCCCTTTCTTTTCAATGAT 2220
QY 2221 GTTGCACACCTTCGCGCATCCAGTGTCTGCTGATTTGCGGACTAGATCAAGAAATATTTAT 2280
Db 2221 GTTGCACACCTTCGCGCATCCAGTGTCTGCTGATTTGCGGACTAGATCAAGAAATATTTAT 2280
QY 2281 TCCAAATGGAAGAACATGATGAGATGTTTGGATGAATCCAGAACCTTGGATATGCTCA 2340
Db 2281 TCCAAATGGAAGAACATGATGAGATGTTTGGATGAATCCAGAACCTTGGATATGCTCA 2340
QY 2341 GGTGTTTCTTCACTTACTGCTGCTGACATGGAGAGTCTTTCGACGATTTACATCCATT 2400
Db 2341 GGTGTTTCTTCACTTACTGCTGCTGACATGGAGAGTCTTTCGACGATTTACATCCATT 2400
QY 2401 TAATAACAGCAGATCCACCAACAGATTAATTCATCTGAGGGAATGTGGTGTTCGAAA 2460
Db 2401 TAATAACAGCAGATCCACCAACAGATTAATTCATCTGAGGGAATGTGGTGTTCGAAA 2460
QY 2461 AAATCCATGCTTGCATCTGCTGATATTCGTTGTTTAAAGGCTGTAATCTTCGAGCGCT 2520
Db 2461 AAATCCATGCTTGCATCTGCTGATATTCGTTGTTTAAAGGCTGTAATCTTCGAGCGCT 2520
QY 2521 GCACACAGTGTAGATGTTGTTGATTCCTCCAGAAAGAAAGACCTCATCCGAAATGA 2580
Db 2521 GCACACAGTGTAGATGTTGTTGATTCCTCCAGAAAGAAAGACCTCATCCGAAATGA 2580
QY 2581 ATGTTCTGGAGTGTGATGTTGGGATATCTACTTGTGTTGTTGGATCAACACATGAT 2640
Db 2581 ATGTTCTGGAGTGTGATGTTGGGATATCTACTTGTGTTGTTGGATCAACACATGAT 2640
QY 2641 CCCGCAAGGCAAGTCCAGCGATGGAATATCCCTCCAGCACCAGCAGATATGCGGACCA 2700
Db 2641 CCCGCAAGGCAAGTCCAGCGATGGAATATCCCTCCAGCACCAGCAGATATGCGGACCA 2700
QY 2701 TGATCTCAATTCAGGAGTGAAGAGTACTTCCACCACTATATTTGTAATCAGAGTTT 2760
Db 2701 TGATCTCAATTCAGGAGTGAAGAGTACTTCCACCACTATATTTGTAATCAGAGTTT 2760
QY 2761 GGAATATAGCAATGCCATGCTGTTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
Db 2761 GGAATATAGCAATGCCATGCTGTTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
QY 2821 TGATCCATGCAAAAGTGTGCTGAGCTTTTCAATTCGAGTGAAGTTCCTCAAGAGCTG 2880
Db 2821 TGATCCATGCAAAAGTGTGCTGAGCTTTTCAATTCGAGTGAAGTTCCTCAAGAGCTG 2880
QY 2881 TGTTCCTCCCTGAAATACCATCTCAGTTGCGGCTTAAAGATACCCAGACTTCATGATAA 2940
Db 2881 TGTTCCTCCCTGAAATACCATCTCAGTTGCGGCTTAAAGATACCCAGACTTCATGATAA 2940
QY 2941 GCGGCAAGCAGCAGTATCTCAGAAAGAGTATTGGAAGCTTTTCAGGAAAGTGA 3000
Db 2941 GCGGCAAGCAGCAGTATCTCAGAAAGAGTATTGGAAGCTTTTCAGGAAAGTGA 3000
QY 3001 GGACAAAGCAGCTCAGGCTAGCTATCGGACCTTCCAGAGAGATGTTGCAAGGAGATC 3060
Db 3001 GGACAAAGCAGCTCAGGCTAGCTATCGGACCTTCCAGAGAGATGTTGCAAGGAGATC 3060
QY 3061 ATATGATGCTATGAGGAGTGTGATGATTTGAGATTCATTCAGCAAGCTTTTGACTA 3120
Db 3061 ATATGATGCTATGAGGAGTGTGATGATTTGAGATTCATTCAGCAAGCTTTTGACTA 3120
QY 3121 CAAACTGGAATATGACCAACAGCTGGGTAAATTAATGAGTACTATGGCATAAAACAGA 3180
Db 3121 CAAACTGGAATATGACCAACAGCTGGGTAAATTAATGAGTACTATGGCATAAAACAGA 3180
QY 3181 GGTGGAATATCTAGTGTGCTATGAGGAGTATGAGGAGTCAAAACCTTTGACCGGAGAAAGA 3240
Db 3181 GGTGGAATATCTAGTGTGCTATGAGGAGTATGAGGAGTCAAAACCTTTGACCGGAGAAAGA 3240

QY 3241 TGCTGAGGCCATTAAGTGTGCTGCTGAGGCGCTTGAGAAAGGAGGAGAGCTGGTTCAA 3300
Db 3241 TGCTGAGGCCATTAAGTGTGCTGCTGAGGCGCTTGAGAAAGGAGGAGAGCTGGTTCAA 3300
QY 3301 GAGCGGTAATGATATAGATGACATGTTTACAAAGGCTTGGCTTGTGACACGTTACATA 3360
Db 3301 GAGCGGTAATGATATAGATGACATGTTTACAAAGGCTTGGCTTGTGACACGTTACATA 3360
QY 3361 TCATCCATACATATTTGGGTTGCTTACATCAGGCGTTGAAAGAGCTCATTTTCAATGCTT 3420
Db 3361 TCATCCATACATATTTGGGTTGCTTACATCAGGCGTTGAAAGAGCTCATTTTCAATGCTT 3420
QY 3421 TCCCTGCTGTTTATGACCAAGCTTAATCCAGATTAAGAAGGCAAAAGCAGTAAACAGGCC 3480
Db 3421 TCCCTGCTGTTTATGACCAAGCTTAATCCAGATTAAGAAGGCAAAAGCAGTAAACAGGCC 3480
QY 3481 AGTTCTCAACTGTGCTCTCTCAGGCGCTCAACTGAGTCACAGATTAGTGTGAAATGAGA 3540
Db 3481 AGTTCTCAACTGTGCTCTCTCAGGCGCTCAACTGAGTCACAGATTAGTGTGAAATGAGA 3540
QY 3541 TTCCAGTCGAGCGTAAAGCTGATATATATTAATAGGCGTGTATCATAGAAAGAAC 3600
Db 3541 TTCCAGTCGAGCGTAAAGCTGATATATATTAATAGGCGTGTATCATAGAAAGAAC 3600
QY 3601 TGTATGCTATGTTGACTACTTTTGTCTTAAACTGCAATGAGCTGCAACATATATGC 3660
Db 3601 TGTATGCTATGTTGACTACTTTTGTCTTAAACTGCAATGAGCTGCAACATATATGC 3660
QY 3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTATATGATGTGATTTGATTTGATC 3720
Db 3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTATATGATGTGATTTGATTTGATC 3720
QY 3721 TTTCTTCTAAA 3731
Db 3721 TTTCTTCTAAA 3731

RESULT 2
US-08-100-874-1/c
; Sequence 1, Application US/08100874
; Patent No. 5498533
; GENERAL INFORMATION:
; APPLICANT: Poovaiyah, B. W.
; APPLICANT: Takezawa, D.
; APPLICANT: Han, T. J.
; APPLICANT: An, G. H.
; TITLE OF INVENTION: Control of Growth and Development of
; TITLE OF INVENTION: Potato Plants
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08100, 874
; FILING DATE: July 30, 1993
; CLASSIFICATION: /800
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REFERENCE/DOCKET NUMBER: 7555-00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (313) 641-1600
; TELEFAX: (313) 641-0270
; TELEX: 287637 Harness UR

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:33:33 ; Search time 142 Seconds
(without alignments)
1616.454 Million cell updates/sec

Title: US-09-782-874-2
Perfect score: 1114
Sequence: 1 MGKTIQVGFPPYLLSAEVRK.....RPVNLSSLRQLSHRLVLR 1114

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database :

- SPTREMBL_21.*
- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1114	100.0	1114	10	Q9ZR58
2	42	3.8	1116	10	Q9ZRY7
3	38	3.4	775	10	Q9ZRX3
4	23	2.1	513	10	Q9ZRX5
5	23	2.1	1107	10	Q9LQV2
6	18	1.6	107	10	Q9ZRX4
7	14	1.3	1133	10	Q82504
8	12	1.1	1196	10	Q9SG02
9	12	1.1	1196	10	Q9LKP0
10	10	0.9	1377	5	Q95WU3
11	9	0.8	583	17	Q26600
12	9	0.8	1069	5	Q9XYP1
13	9	0.8	1638	3	Q9C162
14	8	0.7	198	12	Q39857
15	8	0.7	265	3	Q12044
16	8	0.7	286	16	Q92Y23

17	8	0.7	352	16	Q97ES6
18	8	0.7	444	5	Q95ZS5
19	8	0.7	490	11	Q9CU85
20	8	0.7	515	10	Q9SP43
21	8	0.7	583	13	Q9DG87
22	8	0.7	583	13	Q9IAT4
23	8	0.7	598	4	Q9Y232
24	8	0.7	635	17	Q96YH3
25	8	0.7	830	10	Q65482
26	8	0.7	864	5	Q96064
27	8	0.7	1101	16	Q53347
28	8	0.7	1226	4	Q96S73
29	8	0.7	1283	4	Q96S72
30	8	0.7	1283	4	Q96S71
31	8	0.7	1344	10	Q8S921
32	8	0.7	1564	5	Q9XW49
33	8	0.7	1883	5	Q95ZG7
34	8	0.7	1964	5	Q95ZG6
35	8	0.7	3229	2	Q9ETK0
36	7	0.6	29	16	Q8Z6B7
37	7	0.6	36	6	Q9GKT9
38	7	0.6	43	4	Q9HB48
39	7	0.6	66	16	Q8YNP2
40	7	0.6	70	5	Q9GUX8
41	7	0.6	72	5	Q9NAV2
42	7	0.6	72	5	Q9NAU5
43	7	0.6	72	5	Q9NAU4
44	7	0.6	72	5	Q9NAU3
45	7	0.6	72	5	Q9NAU2
46	7	0.6	72	5	Q9NAT7
47	7	0.6	78	6	Q9RTV1
48	7	0.6	79	16	Q8Z1J9
49	7	0.6	89	4	Q9NT43
50	7	0.6	104	16	Q98111

ALIGNMENTS

RESULT 1

Q9ZR58	ID	Q9ZR58	PRELIMINARY;	PRT;	1114 AA.
AC	Q9ZR58:				
DT	01-MAY-1999	(TRENBLrel. 10, Created)			
DT	01-MAY-1999	(TRENBLrel. 10, Last sequence update)			
DT	01-JUN-2002	(TRENBLrel. 21, Last annotation update)			
DE	RNA-directed RNA polymerase (EC 2.7.7.48).				
GN	RDRP.				
OS	Lycopodium esculentum (Tomato).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Asteridae; euasterids I; Solanales; Solanaceae; Solanum.				
OX	NCBI_TaxID=4081;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=RUTGERS;				
RX	MEDLINE=99055198; PubMed=9836747;				
RA	Schielbel W., Pellissier T., Riedel L., Thalmair S., Schielbel R.,				
RA	Kempe D., Lottspeich F., Sanger H.L., Wassenegger M.;				
RT	"Isolation of an RNA-directed RNA polymerase-specific cDNA clone from				
RT	tomato."				
RL	Plant Cell 10:2087-2102(1998).				
DR	EMBL: Y10403; CAA71421.1;				
DR	InterPro: IPR000504; RNA_rec_mot.				
DR	InterPro: IPR003955; RRM_2.				
DR	Pfam: PF00076; rrm; 1.				
DR	SMART: SM00362; RRM_2; 1.				
KW	Nucleotidyltransferase; RNA-directed RNA polymerase; Transferase.				
SQ	SEQUENCE 1114 AA; 126807 MW; B6D30ACD41DB37FF CRC64;				

Query Match 100.0%; Score 1114; DB 10; Length 1114;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MGKTIQVGFPPYLLSAEVAWSFLEKTYGTGTCVCALEVKQSKGGSRAFAKQVQFADNISADK 60
DB 1 MGKTIQVGFPPYLLSAEVAWSFLEKTYGTGTCVCALEVKQSKGGSRAFAKQVQFADNISADK 60

QY 61 IITLANNRLVFGSSYLKAWEMKTDIVOLRAYVDMQDGTITLNFQCOISDDKFAVLGSTEVS 120
DB 61 IITLANNRLVFGSSYLKAWEMKTDIVOLRAYVDMQDGTITLNFQCOISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKKFFFTLSSGSAADYKQLQSYENITWQVVLHRPYGQNAQFLLIQLFGAPRIYKRL 180
DB 121 IQFGIGLKKFFFTLSSGSAADYKQLQSYENITWQVVLHRPYGQNAQFLLIQLFGAPRIYKRL 180

QY 181 ENSCYSFKEKTPDDQWTRTDFPPSWITGLSSSLCLOFRRCVRPLPNFEESFFHYAERENNI 240
DB 181 ENSCYSFKEKTPDDQWTRTDFPPSWITGLSSSLCLOFRRCVRPLPNFEESFFHYAERENNI 240

QY 241 TLOGFTEFFYSOKSALPNVQPPGEGISIPYKILFKTSSLVHOGCIPGAPALNVFFFLVDP 300
DB 241 TLOGFTEFFYSOKSALPNVQPPGEGISIPYKILFKTSSLVHOGCIPGAPALNVFFFLVDP 300

QY 301 RRRNVACIEHALEKLYIKKECCYDPVWLTEQYDGYLKGROPKSPSITLDDGLVYVRRV 360
DB 301 RRRNVACIEHALEKLYIKKECCYDPVWLTEQYDGYLKGROPKSPSITLDDGLVYVRRV 360

QY 361 LVTCKVYFCGPEVNVSNRVLNRYSEIDNLFVRSFVDEWEKLYSTDLLPKASTGSGVR 420
DB 361 LVTCKVYFCGPEVNVSNRVLNRYSEIDNLFVRSFVDEWEKLYSTDLLPKASTGSGVR 420

QY 421 TNIYERILSTLRKGFVIGDKKFEFLAFSSQLRDNYSVMFASRPGLTANDIRAWMGDFSQ 480
DB 421 TNIYERILSTLRKGFVIGDKKFEFLAFSSQLRDNYSVMFASRPGLTANDIRAWMGDFSQ 480

QY 481 IKNVAKAARLQSGFSGSRETSLVLRHEIEVDPVKVHGTSYVFSOGIGKISGDFAHRYA 540
DB 481 IKNVAKAARLQSGFSGSRETSLVLRHEIEVDPVKVHGTSYVFSOGIGKISGDFAHRYA 540

QY 541 SKCGLQTPSAFAQRYGKYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC 600
DB 541 SKCGLQTPSAFAQRYGKYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC 600

QY 601 YLNQLTLTLSTLGVKEVLEQKQEAVDQDLAILHDSLKAQEALEMLSGENTNITLKA 660
DB 601 YLNQLTLTLSTLGVKEVLEQKQEAVDQDLAILHDSLKAQEALEMLSGENTNITLKA 660

QY 661 LNCYKPDAPFELSMMLQTFRASKLLDLRTSRIFINGRMTMGCLDESRTLEYGVFVQ 720
DB 661 LNCYKPDAPFELSMMLQTFRASKLLDLRTSRIFINGRMTMGCLDESRTLEYGVFVQ 720

QY 721 FTGAGHGEFSDLLHPFNNSRSTNSFILKGNVYVAKNPCLHPGDIRVLKAVNVRAALHHV 780
DB 721 FTGAGHGEFSDLLHPFNNSRSTNSFILKGNVYVAKNPCLHPGDIRVLKAVNVRAALHHV 780

QY 781 DCVVPQKGRPHNEGSGSLDGDIVFCWQDMIPPRQVOPMEYPPAPSIQLDHDVTI 840
DB 781 DCVVPQKGRPHNEGSGSLDGDIVFCWQDMIPPRQVOPMEYPPAPSIQLDHDVTI 840

QY 841 EEVEEYFTNYIVNDSLGITIAHNVVAFADREFDMAMSDPCKKLAEFLSIAYDFPKTGPAE 900
DB 841 EEVEEYFTNYIVNDSLGITIAHNVVAFADREFDMAMSDPCKKLAEFLSIAYDFPKTGPAE 900

QY 901 IPSQLRPKEYDFMDKPKDTSYISERVIGKLRVKVKQKAPQASSIAFTROVAPRSYDAD 960
DB 901 IPSQLRPKEYDFMDKPKDTSYISERVIGKLRVKVKQKAPQASSIAFTROVAPRSYDAD 960

QY 961 MEVDGFEYIDEAFDYKTEYDNKLGNLMDYGIKTEAELLSSGGIMKASKTFDRRKDAEAI 1020
DB 961 MEVDGFEYIDEAFDYKTEYDNKLGNLMDYGIKTEAELLSSGGIMKASKTFDRRKDAEAI 1020

QY 1021 SVAVRALRKEARAFKRRNDIDMLPRASAWHYHTYHTYWGVCYNQGLKRAHFIISFPWCY 1080
DB 1021 SVAVRALRKEARAFKRRNDIDMLPRASAWHYHTYHTYWGVCYNQGLKRAHFIISFPWCY 1080
```

```
QY 1081 YDQLIOIKKDKARNRPVNLSSLRAQLSHRLVLK 1114
DB 1081 YDQLIOIKKDKARNRPVNLSSLRAQLSHRLVLK 1114

RESULT 2
Q9ZRY7 PRELIMINARY: PRT: 1116 AA.
AC Q9ZRY7;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE RNA-directed RNA polymerase.
GN RDRP.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. PETIT HAVANA SRI; TISSUE=LEAF;
RA Schiebel W., Pellissier T., Riedel L., Thalmeir S., Schiebel R.,
RA Kempe D., Lottspeich F., Saenger H.L., Wassenegeger M.;
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT Tomato.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011576; CAA09697.1; .
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1116 AA; 127553 MW; 8D9B89DA826D31DB CRC64;

Query Match 3.8%; Score 42; DB 10; Length 1116;
Best Local Similarity 100.0%; Pred. No. 2.1e-36;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 979 EYDNKLGNLMDYGIKTEAELLSSGGIMKASKTFDRRKDAEAI 1020
DB 981 EYDNKLGNLMDYGIKTEAELLSSGGIMKASKTFDRRKDAEAI 1022

RESULT 3
Q9ZRX3 PRELIMINARY: PRT: 775 AA.
AC Q9ZRX3;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE RNA-directed RNA polymerase (fragment).
GN RDRP.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Schiebel W., Pellissier T., Riedel L., Thalmeir S., Schiebel R.,
RA Kempe D., Lottspeich F., Saenger H.L., Wassenegeger M.;
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT Tomato.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011979; CAA09896.1; .
KW RNA-directed RNA polymerase.
FT NON_TER 1
FT NON_TER 775
SQ SEQUENCE 775 AA; 88778 MW; DE43FE50C6A45B93 CRC64;

Query Match 3.4%; Score 38; DB 10; Length 775;
Best Local Similarity 100.0%; Pred. No. 3.9e-32;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 775 ALHMYDVCVFFQKGRPHNEGSGSLDGDIVFCWD 812
```

Db 468 ALHMVDCVFPQKGRPHNECSGSLDGDIFYVCWD 505
|||||

RESULT 4

Q92RX5 ID Q92RX5 PRELIMINARY; PRT; 513 AA.

AC Q92RX5; 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE RNA-directed RNA polymerase (Fragment).

GN RDRP

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, COLUMBIA; TISSUE=LEAF;

RA Schiebel W., Pelissier T., Riedel L., Thalmeir S., Schiebel R.,

RA Kempe D., Lottspeich F., Saenger H.L., Wassenegeger M.;

RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from

RT Tomato.";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ011977; CAA09894.1;

KW RNA-directed RNA polymerase.

FT NON_TER 1

FT NON_TER 513

SQ SEQUENCE 513 AA; 57802 MW; 924FDA2B70EAD0E CRC64;

Query Match

Best Local Similarity 2.1%; Score 23; DB 10; Length 513;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 RHPNECSGSLDGDIFYVCWDQ 813

|||||

Db 309 RHPNECSGSLDGDIFYVCWDQ 331

RESULT 5

Q9LOV2 ID Q9LOV2 PRELIMINARY; PRT; 1107 AA.

AC Q9LOV2; 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE F10B6.19.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,

RA Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,

RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,

RA Lutos S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,

RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome

RT I.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC006917; AAF79241.1;

SQ SEQUENCE 1107 AA; 126214 MW; F8FF97691FF82187 CRC64;

Query Match

Best Local Similarity 2.1%; Score 23; DB 10; Length 1107;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 RHPNECSGSLDGDIFYVCWDQ 813

|||||

Db 787 RHPNECSGSLDGDIFYVCWDQ 809

RESULT 6

Q92RX4 ID Q92RX4 PRELIMINARY; PRT; 107 AA.

AC Q92RX4;

DT 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE RNA-directed RNA polymerase (Fragment).

GN RDRP

OS Triticum aestivum (Wheat).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

OC Triticeae; Triticum.

OX NCBI_TaxID=4565;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LEAF;

RA Schiebel W., Pelissier T., Riedel L., Thalmeir S., Schiebel R.,

RA Kempe D., Lottspeich F., Saenger H.L., Wassenegeger M.;

RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from

RT Tomato.";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ011978; CAA09895.1;

KW RNA-directed RNA polymerase.

FT NON_TER 1

FT NON_TER 107

SQ SEQUENCE 107 AA; 11908 MW; D74821910CEA964E CRC64;

Query Match

Best Local Similarity 1.6%; Score 18; DB 10; Length 107;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 RHPNECSGSLDGDIFYV 809

|||||

Db 1 RHPNECSGSLDGDIFYV 18

RESULT 7

O82504 ID O82504 PRELIMINARY; PRT; 1133 AA.

AC O82504;

DT 01-NOV-1998 (Tremblrel. 08, Created)

DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)

DE F2P3.11 protein (Putative RNA-directed RNA polymerase).

GN F2P3.11 OR T22B4.110 OR AT4G11130.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA WASHU;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Strong C., Graves T., Duckels G.;
RT "The sequence of A. thaliana F2P3.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[6]
RP SEQUENCE FROM N.A.
RA Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF080120; AAC35535.1; -
DR EMBL; AL049876; CAB43048.1; -
DR EMBL; AL161531; CAB61214.1; -
KW Nucleotidyltransferase; RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 1133 AA; 19323 MW; 32B72CE429B50B9 CRC64;

Query Match 1.3%; Score 14; DB 10; Length 1133;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 PSAFOIRYGGYKGV 562
Db 578 PSAFOIRYGGYKGV 591

RESULT 8
Q9SG02 PRELIMINARY; PRT; 1196 AA.
AC Q9SG02;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Putative RNA-directed RNA polymerase.
GN T1G12.20 OR T9C5.95.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome 1 BAC T1G12 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,

```

```

RA Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20306668; PubMed=10850495;
RA Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C.,
RA Morel J.-B., Jovette D., Lacombe A.-M., Nikic S., Picault N.,
RA Remoue K., Sanial M., Vo T.-A., Vaucheret H.;
RT "Arabidopsis SGS2 and SGS3 genes are required for Posttranscriptional
RT Gene Silencing and Natural Virus Resistance.";
RL Cell 101:533-542(2000).
DR EMBL; AC012329; AAG52184.1; -
DR EMBL; AL132964; CAB71285.1; -
DR EMBL; AF239718; AAF73959.1; -
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136927 MW; 812DEB9CEDC82C8F CRC64;

Query Match 1.1%; Score 12; DB 10; Length 1196;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 KASAWYHYTYHP 1058
Db 1124 KASAWYHYTYHP 1135

RESULT 9
Q9LKP0 PRELIMINARY; PRT; 1196 AA.
AC Q9LKP0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE RNA-dependent RNA polymerase.
GN SDE1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20306669; PubMed=10850496;
RA Dalmay T., Hamilton A., Rudd S., Angell S., Baulcombe D.C.;
RT "An RNA-dependent RNA polymerase gene in Arabidopsis is required for
RT posttranscriptional gene silencing mediated by a transgene but not by
RT a virus.";
RL Cell 101:543-553(2000).
DR EMBL; AF268093; AAF74208.1; -
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136747 MW; C1B1961CDC11F708 CRC64;

Query Match 1.1%; Score 12; DB 10; Length 1196;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 KASAWYHYTYHP 1058
Db 1124 KASAWYHYTYHP 1135

RESULT 10
Q95WU3 PRELIMINARY; PRT; 1377 AA.
AC Q95WU3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE RNA-dependent RNA-polymerase.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;

```

```

RN SEQUENCE FROM N.A.
RA Lujan H.D.;
RT "Identification of an RNA-dependent RNA-polymerase gene in giardia
   intestinalis.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF293414; AAK97084.1; -.
DR InterPro; IPR001865; Ribosomal_S2.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; UNKNOWN.1.
SQ SEQUENCE 1377 AA; 155267 MW; 7453D8646261A0B9 CRC64;

Query Match          0.94; Score 10; DB 5; Length 1377;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 798 SGSDLDDGIY 807
Db 680 SGSDLDDGIY 689

RESULT 11
ID O26600 PRELIMINARY; PRT; 583 AA.
AC O26600;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical protein MTH500.
GN MTH500.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
   Aldredge T., Bashiraden R., Blakely D., Cook R., Gilbert K.,
   Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
   Ji Wani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
   Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
   deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000833; AAB85006.1; -.
DR InterPro; IPR001365; A/AMP_deaminase.
DR PROSITE; PS00485; A_DEAMINASE; UNKNOWN.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 583 AA; 66088 MW; A267B146F0EA6B7A CRC64;

Query Match          0.8%; Score 9; DB 17; Length 583;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 345 SPSITLDDG 353
Db 221 SPSITLDDG 229

RESULT 12
Q9XYP1 PRELIMINARY; PRT; 1069 AA.
AC Q9XYP1;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE Dosa protein.
GN DOSA.
OS Dictyostellum discoideum (Slime mold); Dictyosteliida; Dictyostelium.
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

```

```

OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA Postlethwait P., Steck T.L.;
RT "Dosa, A Gene Related To Hypo-Osmotic Stress Response In Dictyostellum
   discoideum.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF117611; AAD29638.1; -.
SQ SEQUENCE 1069 AA; 122832 MW; BB6815D3B69A1C1C CRC64;

Query Match          0.8%; Score 9; DB 5; Length 1069;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 797 CSGSLLDGD 805
Db 274 CSGSLLDGD 282

RESULT 13
Q9C162 PRELIMINARY; PRT; 1638 AA.
AC Q9C162;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Suppressor of ascus dominance.
GN SAD-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Shiu P.K.T., Raju N.B., Metzberg R.L.;
RT "Meiotic Silencing of Unpaired DNA.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029284; AAK31733.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
SQ SEQUENCE 1638 AA; 184233 MW; 7BC58EA66373D9BE CRC64;

Query Match          0.8%; Score 9; DB 3; Length 1638;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 485 AKYAARLQ 493
Db 653 AKYAARLQ 661

RESULT 14
Q39857 PRELIMINARY; PRT; 198 AA.
AC Q39857;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Coat protein.
OS Grapevine leafroll-associated virus 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=64003;
RN [1]
RP SEQUENCE FROM N.A.
RA Abou-Chanem N.;
RT "The nucleotide sequence of the 3' terminal region of grapevine
   leafroll associated closterovirus 2.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y14131; CAA74566.1; -.
DR InterPro; IPR002679; Closter_coat.
DR Pfam; PF01785; Closter_coat; 1.
SQ SEQUENCE 198 AA; 21630 MW; 15D631E4088D3F01 CRC64;

```

Query Match 0.7%; Score 8; DB 12; Length 198;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 LLSAEVVK 20
| | | | | | | |
Db 27 LLSAEVVK 34

RESULT 15
Q12044
ID Q12044 PRELIMINARY; PRT; 265 AA.
AC Q12044;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Chromosome XV reading frame ORF YOR220W.
GN YOR220W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Boyer J., Fairhead C., Gaillon L., Galisson F., Michaux G.,
RA Thiery A., Dujon B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MOPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FY1679;
RA Galisson F., Dujon B.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: 275128; CAA99438.1; -;
DR EMBL: X92441; CAA63183.1; -;
DR SGD: S0005746; YOR220W.
SQ SEQUENCE 265 AA; 29255 MW; 4A184C5C1218A617 CRC64;

Query Match 0.7%; Score 8; DB 3; Length 265;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 PPKSPSIT 349
| | | | | | | |
Db 252 PPKSPSIT 259

RESULT 16
Q92Y23
ID Q92Y23 PRELIMINARY; PRT; 286 AA.
AC Q92Y23;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein RA1065.
GN RA1065 OR SWA1945.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,

RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.",
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL: AE007292; AAK65723.1; -;
DR InterPro: IPR00620; DUF6.
DR Pfam: PF00892; DUF6; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 286 AA; 30992 MW; 4D8CC747B026FDA4 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 286;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 SKLLDLRT 690
| | | | | | | |
Db 251 SKLLDLRT 258

RESULT 17
Q97ES6
ID Q97ES6 PRELIMINARY; PRT; 352 AA.
AC Q97ES6;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Histidinol-phosphate aminotransferase.
GN CAC3031.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1468;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.I., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007800; AAK80971.1; -;
DR InterPro: IPR002106; AATRNA_ligaseI.
DR InterPro: IPR004839; Aminotransf1/2.
DR InterPro: IPR001917; NHtransf_2.
DR Pfam: PF00155; aminotran_1_2; 1.
DR TIGRfams: TIGR01141; hisC; 1.
DR PROSITE: PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
KW Transferase; Aminotransferase; Complete proteome.
SQ SEQUENCE 352 AA; 40220 MW; 3375021EBEC6BC2E CRC64;

Query Match 0.7%; Score 8; DB 16; Length 352;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 IDNFLRVS 395
| | | | | | | |
Db 326 IDNFLRVS 333

RESULT 18
Q95ZS5
ID Q95ZS5 PRELIMINARY; PRT; 444 AA.
AC Q95ZS5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE F56A8.3a protein.
GN F56A8.3A.
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OX Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z83230; CAC42304.1; -;
DR InterPro; IPR001611; LRR.
DR Pfam; PF00560; LRR; 3.
SQ SEQUENCE 444 AA; 49409 MW; 72E90B3851DA45DB CRC64;

Query Match 0.7%; Score 8; DB 5; Length 444;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 DSLKAEQ 644
Db 156 DSLKAEQ 163

RESULT 19
Q9CUB5 PRELIMINARY; PRT; 490 AA.
AC Q9CUB5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 543014H02Rik protein (Fragment).
GN PEX1 OR 543014H02Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Marsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK017309; BAB30684.1; -;
DR MGD; MGI:1918632; PEX1.
DR InterPro; IPR003593; AAA_Atpase.
DR InterPro; IPR003959; AAA_Atpase_cent.
DR InterPro; IPR003960; AAA_sub.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00674; AAA; UNKNOWN_1.

KW ATP-binding. 1
FT NON_TER 1
SQ SEQUENCE 490 AA; 53417 MW; 01DDC00B35C5BDA6 CRC64;

Query Match 0.7%; Score 8; DB 11; Length 490;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLFGSSY 75
Db 334 RLFGSSY 341

RESULT 20
Q9SP43 PRELIMINARY; PRT; 515 AA.
AC Q9SP43;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 20, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
DE synthase) (ADP-glucose pyrophosphorylase).
GN AGPS.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21210979; PubMed=11311948;
RA Kim I.J., Noh S.J., Lee B.H., Jo J., Kim Y.S., Chung W.I.;
RT "Molecular characterization of cDNA clones for ADP-glucose
RT pyrophosphorylase from Citrus.";
RL Biochim. Biophys. Acta 1518:324-328(2001).
CC -1- CATALYTIC ACTIVITY: ATP + ALPHA-D-GLUCOSE 1-PHOSPHATE =
CC DIPHOSPHATE + ADP-GLUCOSE.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
DR EMBL; AF184597; AAD56041.1; -;
DR InterPro; IPR001825; NTP_transferase.
DR Pfam; PF00483; NTP_transferase; 1.
DR PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE; PS00810; ADP_GLC_PYROPHOSPH_3; 1.
KW Glycogen biosynthesis; Nucleotidyltransferase; Transferase.
SQ SEQUENCE 515 AA; 56814 MW; 925CED3EC3F44258 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 515;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 LAFSSQL 452
Db 34 LAFSSQL 41

RESULT 21
Q9DG87 PRELIMINARY; PRT; 583 AA.
AC Q9DG87;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Glutamic acid decarboxylase isoform 65 (EC 4.1.1.15).
GN GAD65.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.

```
RC TISSUE-BRAIN, AND PITUITARY;
RX MEDLINE=21103894; PubMed=11181814;
RA Bosma P.T., Blazquez M., Fraser E.J., Schulz R.W., Docherty K.,
RT Trudeau V.L.;
RT "Sex steroid regulation of glutamate decarboxylase mRNA expression in
RT goldfish brain is sexually dimorphic.";
RL J. Neurochem. 76:945-956(2001)
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
CC TYRDC).
DR EMBL; AF045594; AAG33931.1; -.
DR InterPro; IPR002129; Pyridoxal_dec.
DR Pfam; PF00282; Pyridoxal_dec.1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Decarboxylase; Lyase; Pyridoxal phosphate.
SQ SEQUENCE 583 AA; 65375 MW; 8C43E1A264D95001 CRC64;

Query Match 0.7%; Score 8; DB 13; Length 583;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LYSTDLLP 411
DB 90 LYSTDLLP 97

RESULT 22
Q9IAT4
ID Q9IAT4 PRELIMINARY; PRT; 583 AA.
AC Q9IAT4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Glutamic acid decarboxylase isoform 65 (EC 4.1.1.15).
GN GAD65.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Bosma P.T., Blazquez M., Docherty K., Trudeau V.L.;
RT "Cloning and expression of multiple forms of glutamate decarboxylase
RT from goldfish.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
CC TYRDC).
DR EMBL; AF149832; AAF73186.1; -.
DR InterPro; IPR002129; Pyridoxal_dec.
DR Pfam; PF00282; Pyridoxal_dec.1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Decarboxylase; Lyase; Pyridoxal phosphate.
SQ SEQUENCE 583 AA; 65361 MW; 9C4326BFD2023DB6 CRC64;

Query Match 0.7%; Score 8; DB 13; Length 583;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LYSTDLLP 411
DB 90 LYSTDLLP 97

RESULT 23
Q9Y232
ID Q9Y232 PRELIMINARY; PRT; 598 AA.
AC Q9Y232;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
```

```
DE Testis-specific CHROMODOMAIN Y-like protein.
GN CDYL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE=99206615; PubMed=10192397;
RA Lahn B.T., Page D.C.;
RT "Retroposition of autosomal mRNA yielded testis-specific gene family
RT on human Y chromosome.";
RL Nat. Genet. 21:429-433(1999).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: CONTAINS 1 "CHROMO" DOMAIN.
DR EMBL; AF081259; AAD22735.1; -.
DR EMBL; AF081258; AAD22734.1; -.
DR HSSP; P23197; IAP0.
DR InterPro; IPR000953; Chromo.
DR Pfam; PF00385; Chromo; 1.
DR Pfam; PF00378; ECH; 1.
DR PRINTS; PR00504; CHROMODOMAIN.
DR SMART; SM00298; CHROMO; 1.
DR PROSITE; PS00598; CHROMO_1; 1.
DR PROSITE; PS00013; CHROMO_2; 1.
KW Nuclear protein.
SQ SEQUENCE 598 AA; 66419 MW; 503D20E04BDBDC90 CRC64;

Query Match 0.7%; Score 8; DB 4; Length 598;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 SRSTNSNF 746
DB 138 SRSTNSNF 145

RESULT 24
Q96YH3
ID Q96YH3 PRELIMINARY; PRT; 635 AA.
AC Q96YH3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative glutamate synthase large subunit.
GN ST2198.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Karabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sakine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000989; BAB67304.1; -.
DR InterPro; IPR003009; FMN_enzyme.
DR InterPro; IPR002932; Glu_synthase.
DR Pfam; PF01645; Glu_synthase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 635 AA; 68139 MW; 8D6BEE204F66812F CRC64;

Query Match 0.7%; Score 8; DB 17; Length 635;
```

```
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 SRETLSVL 505
DB 319 SRETLSVL 326
|||||
162 LITLSTL 169

RESULT 26
O96064 PRELIMINARY; PRT; 864 AA.
AC O96064;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Paramyosin.
OS Mytilus galloprovincialis (Mediterranean mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=29158;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ANTERIOR BYSSUS RETRACTOR MUSCLE;
RX MEDLINE=20073061; PubMed=10607366;
RA Wakabe S., Iwasaki K., Furubara D., Hirayama Y., Nakaya M.,
RA Kikuchi K.;
RT "Complete amino acid sequence of Mytilus anterior byssus retractor
    paramyosin and its putative phosphorylation site.";
RL J. Exp. Zool. 286:24-35(2000).
DR EMBL; AB016070; BAA36517.1; -
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR004827; TFIIBZIP.
DR Pfam; PF01576; Myosin_tail; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 864 AA; 99572 MW; EF7C8ED73A5742B0 CRC64;

Query Match 0.7%; Score 8; DB 5; Length 864;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1101 SSLRAQLS 1108
DB 280 SSLRAQLS 287
|||||

RESULT 27
O53347 PRELIMINARY; PRT; 1101 AA.
AC O53347;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative ATP-dependent DNA helicase (Helicase, UVID/Rep family).
GN R3201C OR MT014.45C OR MT3295.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsbury T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
    complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
```

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021646; CAA16666.1; -;
 DR EMBL; AE007142; AAK47638.1; -;
 DR HSSP; P56255; 1PJR.
 DR TIGR; MT3235; -;
 DR TubercuList; rv3201c; -;
 DR InterPro; IPR000212; UvrD-helicase.
 DR Pfam; PF00580; UvrD-helicase; 1.
 KW Helicase; Complete proteome.
 SQ SEQUENCE 1101 AA; 116687 MW; 269C6AAD3657412B CRC64;

Query Match 0.7%; Score 8; DB 16; Length 1101;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1021 SVAVRALR 1028
 Db 374 SVAVRALR 381
 |||||

RESULT 28

Q96S73 PRELIMINARY; PRT; 1226 AA.
 AC Q96S73;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Pexlp-634del1690.
 GN PEX1-634DEL1690.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tamura S., Matsumoto N., Imamura A., Shimozawa N., Suzuki Y.,
 RA Kondo N., Fujiki Y.;
 RT "Phenotype-Genotype Relationships in Peroxisome Biogenesis Disorders
 RT of PEX1-Defective Complementation Group 1 Are Defined by Pexlp-Pex6p
 RT Interaction";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB052090; BAB59061.1; -;
 DR InterPro; IPR003959; AAA_ATPase_cent.
 DR InterPro; IPR003960; AAA_sub.
 DR Pfam; PF00004; AAA; 2.
 DR PROSITE; PS00674; AAA; UNKNOWN_1.
 SQ SEQUENCE 1226 AA; 136584 MW; 28125C8B785DA8E2 CRC64;

Query Match 0.7%; Score 8; DB 4; Length 1226;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75
 Db 1070 RLYFGSSY 1077
 |||||

RESULT 29

Q96S72 PRELIMINARY; PRT; 1283 AA.
 AC Q96S72;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Pexlp1L664P.
 GN PEX1L664P.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tamura S., Matsumoto N., Imamura A., Shimozawa N., Suzuki Y.,
 RA Kondo N., Fujiki Y.;
 RT "Phenotype-Genotype Relationships in Peroxisome Biogenesis Disorders
 RT of PEX1-Defective Complementation Group 1 Are Defined by Pexlp-Pex6p
 RT Interaction";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB052091; BAB59062.1; -;
 DR InterPro; IPR003959; AAA_ATPase_cent.
 DR InterPro; IPR003960; AAA_sub.
 DR Pfam; PF00004; AAA; 2.
 DR PROSITE; PS00674; AAA; UNKNOWN_1.
 SQ SEQUENCE 1283 AA; 142850 MW; C33AD8A68893D77A CRC64;

Query Match 0.7%; Score 8; DB 4; Length 1283;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75
 Db 1127 RLYFGSSY 1134
 |||||

RESULT 30

Q96S71 PRELIMINARY; PRT; 1283 AA.
 AC Q96S71;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE PexlpG843D.
 GN PEX1G843D.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tamura S., Matsumoto N., Imamura A., Shimozawa N., Suzuki Y.,
 RA Kondo N., Fujiki Y.;
 RT "Phenotype-Genotype Relationships in Peroxisome Biogenesis Disorders
 RT of PEX1-Defective Complementation Group 1 Are Defined by Pexlp-Pex6p
 RT Interaction";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB052092; BAB59063.1; -;
 DR InterPro; IPR003959; AAA_ATPase_cent.
 DR InterPro; IPR003960; AAA_sub.
 DR Pfam; PF00004; AAA; 2.
 DR PROSITE; PS00674; AAA; UNKNOWN_1.
 SQ SEQUENCE 1283 AA; 142924 MW; 9CF230C26F02AFD9 CRC64;

Query Match 0.7%; Score 8; DB 4; Length 1283;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75
 Db 1127 RLYFGSSY 1134
 |||||

RESULT 31

Q9S921 PRELIMINARY; PRT; 1344 AA.
 AC Q9S921;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Chloroplast-resident DNA methyltransferase.
 GN CRMET1.

```

OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CC683;
RX MEDLINE=21980658; PubMed=11983892;
RA Nishiyama R., Itoh M., Yamaguchi Y., Koizumi N., Sano H.;
RT "A chloroplast-resident DNA methyltransferase is responsible for
RT hypermethylation of chloroplast genes in Chlamydomonas maternal
RT gametes.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:5925-5930(2002).
KW EMBL; AB073989; BAB91073.1; --
KW Transferase; Methyltransferase.
SQ SEQUENCE 1344 AA; 142361 MW; 96D3876E7EEB477F CRC64;

Query Match 0.7%; Score 8; DB 10; Length 1344;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1100 LSSLRAQL 1107
DB 150 LSSLRAQL 157
|||||||

RESULT 32
Q9XW49 PRELIMINARY; PRT; 1564 AA.
AC Q9XW49;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Y53C10A.9 protein.
DE Y53C10A.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL033536; CAA22142.1; --
DR InterPro; IPR003439; ABC.transportr.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00005; ABC.trans; 2.
DR ProDom; PD000006; ABC.transportr; 2.
DR PROSITE; PS00211; ABC.TRANSPORTER; UNKNOWN_2.
DR PROSITE; PS00636; DNAJ_1; UNKNOWN_1.
SQ SEQUENCE 1564 AA; 176701 MW; DC66A3E7ED8DC50F CRC64;

Query Match 0.7%; Score 8; DB 5; Length 1564;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 606 LITLSTL 613
DB 254 LITLSTL 261
|||||||

RESULT 33
Q95ZG7 PRELIMINARY; PRT; 1883 AA.
AC Q95ZG7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)

```

```

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative RNA dependent RNA polymerase (Fragment).
GN RPPA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RA Nellen W.;
RT "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
RT and dsRNase.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314909; CAC41974.1; --
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001201; PAP_25A_core.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW ATP-binding; Helicase.
FT NON_TER 1883
SQ SEQUENCE 1883 AA; 217454 MW; 1B1D9B9738B15094 CRC64;

Query Match 0.7%; Score 8; DB 5; Length 1883;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 SGSDLDGD 805
DB 1353 SGSDLDGD 1360
|||||||

RESULT 34
Q95ZG6 PRELIMINARY; PRT; 1964 AA.
AC Q95ZG6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative RNA dependent RNA polymerase (Fragment).
GN RPPB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RA Nellen W.;
RT "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
RT and dsRNase.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314910; CAC41975.1; --
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001201; PAP_25A_core.
DR Pfam; PF00271; Helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
SQ SEQUENCE 1964 AA; 227413 MW; B83CA9098675E645 CRC64;

Query Match 0.7%; Score 8; DB 5; Length 1964;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 SGSDLDGD 805
DB 1177 SGSDLDGD 1184
|||||||

```

```

RESULT 35
Q9ETK0
ID Q9ETK0 PRELIMINARY; PRT; 3229 AA.
AC Q9ETK0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Putative methylase (or helicase).
GN ORF26.
OS Corynebacterium equi (Rhodococcus equi).
OG Plasmid PREAT701 (P33701), and Plasmid virulence.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=43767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC33701; PLASMID-PREAT701 (P33701);
RA Takai S., Sekizaki T., Kakuda T., Nakamura M., Suzuki K., Ogino N.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC33701, AND 103; PLASMID-PREAT701 (P33701), AND VIRULENCE;
RX PubMed11083803;
RA Takai S., Hines S.A., Sekizaki T., Nicholson V.M., Alperin D.A.,
RA Osaki M., Takamatsu D., Nakamura M., Suzuki K., Ogino N., Kakuda T.,
RA Dan H., Prescott J.F.;
RT "DNA sequence and comparison of virulence plasmids from Rhodococcus
equi ATCC 33701 and 103."
RL Infect. Immun. 68:6840-6847(2000).
DR EMBL; AP001204; BAB16635.1; -
DR EMBL; AFI16907; AAG21729.1; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
KW ATP-binding; Helicase; Methyltransferase; Plasmid.
SQ SEQUENCE 3229 AA; 351003 MW; CLE03A6050542421A CRC64;

Query Match 0.7%; Score 8; DB 2; Length 3229;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 952 VARRSYDA 959
Db 2704 VARRSYDA 2711

RESULT 36
Q8Z6B7
ID Q8Z6B7 PRELIMINARY; PRT; 29 AA.
AC Q8Z6B7;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein STX1874a.
GN STX1874a.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,

Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi Cm18."
RL Nature 413:848-852(2001).
DR EMBL; AL627271; CAD02107.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 29 AA; 3223 MW; 2A834B159BD06DB4 CRC64;

Query Match 0.6%; Score 7; DB 16; Length 29;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 990 YYGKTE 996
Db 8 YYGKTE 14

RESULT 37
Q9GKJ9
ID Q9GKJ9 PRELIMINARY; PRT; 36 AA.
AC Q9GKJ9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Prostatic acid phosphatase (Fragment).
GN ACPp.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Van Poucke M., Yerie M., Tuggle C., Chardon P., Van Zeveren A.,
RA Peelman L.J.;
RT "Integration of porcine chromosome 13 maps."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229111; AAG41124.1; -
DR HSSP; P15309; 2HPA.
DR InterPro; IPR000560; HisAc.phosphatse.
DR Pfam; PF00328; acid_phosphat; 1.
FT NON_TER 1
FT NON_TER 36
SQ SEQUENCE 36 AA; 3943 MW; 50C57B8C770EAL0C CRC64;

Query Match 0.6%; Score 7; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 PPEGISI 268
Db 12 PPEGISI 18

RESULT 38
Q9HB46
ID Q9HB46 PRELIMINARY; PRT; 43 AA.
AC Q9HB46;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Hypothetical 5.4 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Qihan L., Chenhong D., Hongling Z.;
RT "Putative human lung cell protein."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF281279; AAG13933.1; -
KW Hypothetical protein.

```

SQ SEQUENCE 43 AA; 5388 MW; 0FB612B8E31EBA71 CRC64;
Query Match 0.68; Score 7; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 LKKEFFF 133
|||||||
Db 21 LKKEFFF 27

RESULT 39

ID Q8YNP2 PRELIMINARY; PRT; 66 AA.
AC Q8YNP2;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Hypothetical protein Asr4522.
GN ASR4522.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RI DNA Res. 8:205-213(2001).
RL EMBL; AP003596; BAB76221.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 66 AA; 7573 MW; AB4150A4459CAA9 CRC64;

Query Match 0.68; Score 7; DB 16; Length 66;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 VQPPEGI 266
|||||||
Db 55 VQPPEGI 61

RESULT 40

ID Q9GUX8 PRELIMINARY; PRT; 70 AA.
AC Q9GUX8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE Foot protein-3A.
GN MGFP-3A.
OS Mytilus galloprovincialis (Mediterranean mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=29158;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FOOT;
RL Inoue K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FOOT;
RX MEDLINE=96305382; PubMed=8706704;
RA Inoue K., Takeuchi Y., Miki D., Odo S., Harayama S., Waite H.;
RT "Cloning, sequencing and sites of expression of genes for the
hydroxyarginine-containing adhesive-plaque protein of the mussel
Mytilus galloprovincialis.";

RL Eur. J. Biochem. 239:172-176(1996).
DR EMBL; AB049579; BAB16314.1; -
SQ SEQUENCE 70 AA; 8003 MW; A902B3E300F5B9EA CRC64;

Query Match 0.6%; Score 7; DB 5; Length 70;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 555 RYGGYKG 561
|||||||
Db 49 RYGGYKG 55

RESULT 41

ID Q9NAV2 PRELIMINARY; PRT; 72 AA.
AC Q9NAV2;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Foot protein-3 precursor variant1.
GN MEFP3.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX Warner S.C., Waite J.H.;
RT "Expression of multiple forms of an adhesive plaque protein in an
individual mussel, Mytilus edulis.";
RL Mar. Biol. 134:729-734(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX Warner S.C., Waite J.H.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF286136; AAF89278.1; -
SQ SEQUENCE 72 AA; 8295 MW; 7D42746A910D99 CRC64;

Query Match 0.6%; Score 7; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 555 RYGGYKG 561
|||||||
Db 52 RYGGYKG 58

RESULT 42

ID Q9NAU5 PRELIMINARY; PRT; 72 AA.
AC Q9NAU5;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Foot protein-3 precursor variant10.
GN MEFP3.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX Warner S.C., Waite J.H.;
RT "Expression of multiple forms of an adhesive plaque protein in an
individual mussel, Mytilus edulis.";
RL Mar. Biol. 134:729-734(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX Warner S.C., Waite J.H.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF286147; AAF89289.1; -
SQ SEQUENCE 72 AA; 8214 MW; 7600241D884D6BF4 CRC64;

```
Query Match          0.6%; Score 7; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561
  |||||
Db 52 RYGGYKG 58

RESULT 43
Q9NAU4          PRELIMINARY; PRT; 72 AA.
ID Q9NAU4
AC Q9NAU4
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Foot protein-3 precursor variant11.
GN MEFP3.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RA Warner S.C., Waite J.H.;
RT "Expression of multiple forms of an adhesive plaque protein in an
  individual mussel, Mytilus edulis.";
RL Mar. Biol. 134:729-734(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Warner S.C., Waite J.H.;
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF286148; AAF89290.1;
DR EMBL; AF286148; AAF89290.1;
SQ SEQUENCE 72 AA; 8249 MW; 3198746A89796D81 CRC64;

Query Match          0.6%; Score 7; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561
  |||||
Db 52 RYGGYKG 58

RESULT 44
Q9NAU3          PRELIMINARY; PRT; 72 AA.
ID Q9NAU3
AC Q9NAU3
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Foot protein-3 precursor variant12.
GN MEFP3.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RA Warner S.C., Waite J.H.;
RT "Expression of multiple forms of an adhesive plaque protein in an
  individual mussel, Mytilus edulis.";
RL Mar. Biol. 134:729-734(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Warner S.C., Waite J.H.;
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF286149; AAF89291.1;
DR EMBL; AF286149; AAF89291.1;
SQ SEQUENCE 72 AA; 8298 MW; 7602746A8D6D6B3F4 CRC64;

Query Match          0.6%; Score 7; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561
  |||||
Db 52 RYGGYKG 58

RESULT 45
Q9NAU2          PRELIMINARY; PRT; 72 AA.
ID Q9NAU2
AC Q9NAU2
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Foot protein-3 precursor variant13.
GN MEFP3.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RA Warner S.C., Waite J.H.;
RT "Expression of multiple forms of an adhesive plaque protein in an
  individual mussel, Mytilus edulis.";
RL Mar. Biol. 134:729-734(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Warner S.C., Waite J.H.;
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF286150; AAF89292.1;
DR EMBL; AF286150; AAF89292.1;
SQ SEQUENCE 72 AA; 8261 MW; B2B1746A98989D90 CRC64;

Query Match          0.6%; Score 7; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561
  |||||
Db 52 RYGGYKG 58

RESULT 46
Q9NAU7          PRELIMINARY; PRT; 72 AA.
ID Q9NAU7
AC Q9NAU7
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Foot protein-3 precursor variant19.
GN MEFP3.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RA Warner S.C., Waite J.H.;
RT "Expression of multiple forms of an adhesive plaque protein in an
  individual mussel, Mytilus edulis.";
RL Mar. Biol. 134:729-734(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Warner S.C., Waite J.H.;
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF286156; AAF89298.1;
DR EMBL; AF286156; AAF89298.1;
SQ SEQUENCE 72 AA; 8327 MW; 6CB3DF69ED6D6BF4 CRC64;

Query Match          0.6%; Score 7; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561
  |||||
Db 52 RYGGYKG 58
```


Db 52 RYGGK 58
|||||||

RESULT 47

Q9TTV1 PRELIMINARY; PRT; 78 AA.
AC Q9TTV1
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Breast basic conserved protein 1 (Fragment).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PULMONARY VEIN;
RA Sander F.C., Raj J.U.;
RT "Breast Basic Conserved Gene (BBCL) In Fetal Pulmonary Vascular Development";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF192977; AAF03752.1;
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13E; 1.
DR ProDom; PD004443; Ribosomal_L13E; 1.
FT NON_TER 1 78
FT SEQUENCE 78 AA; 8924 MW; 54EBA54962C35E2D CRC64;
SQ SEQUENCE 78 AA; 8924 MW; 54EBA54962C35E2D CRC64;

Query Match 0.6%; Score 7; DB 6; Length 78;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 VDPRRN 304
|||||||
Db 67 VDPRRN 73

RESULT 48

Q8ZLJ9 PRELIMINARY; PRT; 79 AA.
AC Q8ZLJ9
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE putative membrane protein.
GN STY4565.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18";
RL Nature 413:848-852(2001).
DR EMBL; AJ627282; CAB09341.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 79 AA; 8818 MW; 747F99DE87ECB71C CRC64;

Query Match 0.6%; Score 7; DB 16; Length 79;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 456 SVMFAS 462
|||||||
Db 73 SVMFAS 79

RESULT 49

Q9NT43 PRELIMINARY; PRT; 89 AA.
AC Q9NT43
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 10.1 kDa protein (Fragment).
GN DKFZP434P2119.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Otterwilder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137543; CAB70801.1;
DR InterPro; IPR004934; Tropomodulin.
DR Pfam; PF03250; Tropomodulin; 1.
KW Hypothetical protein.
FT NON_TER 1 89
FT SEQUENCE 89 AA; 10055 MW; 568C62F6D61D39A CRC64;
SQ SEQUENCE 89 AA; 10055 MW; 568C62F6D61D39A CRC64;

Query Match 0.6%; Score 7; DB 4; Length 89;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 652 ENTNLIK 658
|||||||
Db 48 ENTNLIK 54

RESULT 50

Q98I11 PRELIMINARY; PRT; 104 AA.
AC Q98I11
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein ml12618.
GN ML12618.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003000; BAB49705.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 104 AA; 10787 MW; 1743BC06301BAACB CRC64;

Query Match 0.6%; Score 7; DB 16; Length 104;

Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 347 SITLDDG 353
| | | | |
Db 60 SITLDDG 66

Search completed: November 6, 2002, 03:37:49
Job time : 163 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:39:18 ; Search time 45 Seconds
(without alignments)
3298.693 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVGFPPYLLSAEVK.....RPVLMSSLRQLSHRLVVK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5864	100.0	1114	22	AAU10006
2	5864	100.0	1114	22	AAU10006
3	3643	62.1	1125	21	AAU28533
4	3140.5	53.6	1108	21	AAU28530
5	1621	27.6	1196	22	AAU67235
6	1591	27.1	1164	21	AAU28531
7	1430.5	24.4	836	21	AAU28532
8	1182	20.2	218	22	AAU00898
9	1013.5	17.3	417	21	AAU28529
10	697.5	11.9	230	21	AAU28534

11	297	5.1	1402	21	AAU13956
12	205	3.5	60	21	AAU28537
13	197	3.4	90	21	AAU28536
14	167	2.8	111	21	AAU28535
15	146	2.5	2060	23	AAU20967
16	140	2.4	2029	23	ABU68683
17	129.5	2.2	849	23	ABU91193
18	128	2.2	157	21	AAU28528
19	126.5	2.2	2548	21	AAU53193
20	123	2.1	1498	22	ABU05469
21	120	2.0	996	22	ABU30156
22	117.5	2.0	908	22	ABU12986
23	117.5	2.0	1345	22	ABU13923
24	117	2.0	935	22	ABU10172
25	116.5	2.0	544	22	AAU01219
26	116.5	2.0	546	20	AAU06850
27	116.5	2.0	751	23	ABU26865
28	115	2.0	766	21	AAU14283
29	115	2.0	774	20	AAU37454
30	115	2.0	872	23	AAU071619
31	114.5	2.0	544	22	AAU01222
32	114.5	2.0	546	20	AAU06853
33	114.5	2.0	776	23	ABU49921
34	114	1.9	957	18	AAU20841
35	113.5	1.9	831	19	AAU54372
36	113.5	1.9	3226	22	ABU28408
37	113	1.9	763	21	AAU14284
38	112.5	1.9	742	22	ABU22122
39	112.5	1.9	988	19	AAU56107
40	112.5	1.9	1010	22	ABU20134
41	112.5	1.9	1184	22	AAU33733
42	112.5	1.9	1188	22	AAU36532
43	112	1.9	1198	23	ABU5148
44	111.5	1.9	544	22	AAU01225
45	111.5	1.9	546	20	AAU06856

ALIGNMENTS

RESULT 1

AAU10006
ID AAU10006 standard; Protein; 1114 AA.

XX AAU10006;

XX 08-MAY-2002 (first entry)

XX Tomato RNA-directed RNA polymerase (RdRP) protein.

XX Tomato: RdRP: RNA-directed RNA polymerase; in vitro transcription;
XX cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
XX transgenic plant; transgenic animal; cancer; viral infection;
XX immunoprecipitation; immunolocalisation; EC 2.7.7.48; gene therapy.

XX Lycopersicon esculentum.

XX US2001023067-A1.

XX 20-SEP-2001.

XX 08-FEB-2001; 2001US-0782874.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-595798/67.

XX N-PSDB; AAS17837.

XX

New nucleic acid molecule encoding a polypeptide having the enzymatic activity of RNA-directed RNA polymerase, for modulating gene expression and treating cancer and virus infection in human and animals -

Claim 1: Page 19-22; 34pp; English.

This sequence represents the tomato RNA-directed RNA polymerase (RdRP) protein of the invention. The invention comprises the nucleic acid and protein sequences of RdRP. The protein of the invention can catalyze in vitro transcription of short single stranded RNAs into DNA molecules, this transcription can be either primed by RNA or DNA oligonucleotides, or be unprimed. The protein may have cytosstatic or virucide activities. The sequences of the invention may be used in gene therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template nucleic acid molecule derived from a nucleic acid molecule which causes a disease are useful for treating a disease caused by the undesired expression or overexpression of a nucleic acid molecule in a human, rat or mouse, by administering the molecules. This system can be used in the preparation of a pharmaceutical composition and for inhibiting expression of any desired gene by transferring the RdRP system to organisms that either lack a comparable mechanism or do not sufficiently express their own RdRP. An antibody or an antagonist or inhibitor to the protein are useful for inhibiting RNA directed RNA synthesis and for ensuring stable heterologous, gene expression in transgenic organisms. The sequence is useful for probes and/or for the control of gene expression, as primers for amplification of nucleic acid molecules and as tools for the detection of expression of the cDNA molecules. Additionally, nucleotide and protein sequences are useful for suppression of undesired gene expression in humans and animals. The RdRP is useful as a therapeutic agent for the control of cancer and virus infection in humans and animals and the antibody is useful for immunoprecipitation or immunolocalisation of the protein, identification of polypeptides interacting with it and screening expression libraries.

Sequence 1114 AA;

Query Match 100.0%; Score 5864; DB 22; Length 1114;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVFGPPYLLSAEYKVSFLKYGTYGTVCALEYKQSGSRAPAKVQFADNISADK 60
 DB 1 MGKTIQVFGPPYLLSAEYKVSFLKYGTYGTVCALEYKQSGSRAPAKVQFADNISADK 60

QY 61 IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDQMDGITLNFQCIISDDKFAVLGSTEVS 120
 DB 61 IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDQMDGITLNFQCIISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKKTFPFLSSGSADYKLGQSYENIWOVYLHRPYGONAFLLIQLFGAPRIYKRL 180
 DB 121 IQFGIGLKKTFPFLSSGSADYKLGQSYENIWOVYLHRPYGONAFLLIQLFGAPRIYKRL 180

QY 181 ENSCYSFKETPDQWRTTDFPPSWIGLSSLCLOFRGVRPLNPFESFFHYAERENNI 240
 DB 181 ENSCYSFKETPDQWRTTDFPPSWIGLSSLCLOFRGVRPLNPFESFFHYAERENNI 240

QY 241 TLOGTGFTFFVSOKALPVNVPQPGISIPYKILFKISSLVQHGCIPGALNVFFRLVDP 300
 DB 241 TLOGTGFTFFVSOKALPVNVPQPGISIPYKILFKISSLVQHGCIPGALNVFFRLVDP 300

QY 301 RRRNVACIEHALEKLYIYKECCDVPVWLTQVYDGYLKGQPKPSITLDDGLVYVRRY 360
 DB 301 RRRNVACIEHALEKLYIYKECCDVPVWLTQVYDGYLKGQPKPSITLDDGLVYVRRY 360

QY 361 LVTPCKVYFCGPEVNVNRLVNSYEDIDNPLRVSFVDEWEKLYSTDLLPKASTGSVGR 420
 DB 361 LVTPCKVYFCGPEVNVNRLVNSYEDIDNPLRVSFVDEWEKLYSTDLLPKASTGSVGR 420

QY 421 TNYIERILSLRKGFGVIGDKKFEFLAFSSQLRDNVSWMFASRPGLTANDIRAWMGDFSQ 480
 DB 421 TNYIERILSLRKGFGVIGDKKFEFLAFSSQLRDNVSWMFASRPGLTANDIRAWMGDFSQ 480

QY 481 IKNVAKYAARLQSGFSGSRRTLSVLRHEIEVIPDKVKVHGTSYVFSGIGKISGDFAHRYA 540

DB 481 IKNVAKYAARLQSGFSGSRRTLSVLRHEIEVIPDKVKVHGTSYVFSGIGKISGDFAHRYA 540
 QY 541 SKCGLQYTPSAFQIRYRGYGVGVDPDSMKLSLRKSMKSYESDNKILDLVLSWQYQPC 600
 DB 541 SKCGLQYTPSAFQIRYRGYGVGVDPDSMKLSLRKSMKSYESDNKILDLVLSWQYQPC 600

QY 601 YLNRQLITLLSTLGVKDEVLQKQKAVDQLDAILHDSLKAQALELMSFGENTNLKAM 660
 DB 601 YLNRQLITLLSTLGVKDEVLQKQKAVDQLDAILHDSLKAQALELMSFGENTNLKAM 660

QY 661 LNCGYKPDABEPFLSMQLQFRASKLLDLRTSRIFIPNGRTMNGCLDESRITLEYGVQVYQ 720
 DB 661 LNCGYKPDABEPFLSMQLQFRASKLLDLRTSRIFIPNGRTMNGCLDESRITLEYGVQVYQ 720

QY 721 FTGAGHGEFSDLLHPNNSRSTNSNFIKGNVVVAKNCPCLHPGDIRVLKAVNVRALEHHV 780
 DB 721 FTGAGHGEFSDLLHPNNSRSTNSNFIKGNVVVAKNCPCLHPGDIRVLKAVNVRALEHHV 780

QY 781 DCVVFFQKGRPHNPNCSGSDLDGDIYFVCWDQDMIPPRQVOPMEYPPAPSIQLDHDVTI 840
 DB 781 DCVVFFQKGRPHNPNCSGSDLDGDIYFVCWDQDMIPPRQVOPMEYPPAPSIQLDHDVTI 840

QY 841 EEVEEYFTNYVNDLSGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGPAE 900
 DB 841 EEVEEYFTNYVNDLSGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGPAE 900

QY 901 IPSQLPKPEYPMQPKDTSYISERVIGKIPKVKDKAPQASSIATFTRDVARRSYDAD 960
 DB 901 IPSQLPKPEYPMQPKDTSYISERVIGKIPKVKDKAPQASSIATFTRDVARRSYDAD 960

QY 961 MEYDGFEDYIDEAFDYKTEYDNKLGMLNDYIYGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020
 DB 961 MEYDGFEDYIDEAFDYKTEYDNKLGMLNDYIYGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020

QY 1021 SVAVRALRKEARAWFKRRNDIDMLPKASAWHYVTHPTYWGCVNGGLKRAHFISFPWCV 1080
 DB 1021 SVAVRALRKEARAWFKRRNDIDMLPKASAWHYVTHPTYWGCVNGGLKRAHFISFPWCV 1080

QY 1081 YDOLIGIKDKARNRPVNLSSLRAOLSHRLVILK 1114
 DB 1081 YDOLIGIKDKARNRPVNLSSLRAOLSHRLVILK 1114

RESULT 2
 ID AAE00897
 XX AC AAE00897; standard; Protein; 1114 AA.
 XX AC AAE00897;
 XX DT 04-JUL-2001 (first entry)
 XX DE Tomato C-protein having RNA-directed RNA polymerase (RdRP) activity.
 XX KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression; transgenic plant; tissue culture; plant breeding; therapy; C-protein.
 XX OS Lycopersicon esculentum.
 XX FH Key Location/Qualifiers
 FT Region 331..346
 FT /label= P432
 FT /note= "C-protein-specific antibody for immunisation of rabbits"
 FT Region 377..390
 FT /label= P430
 FT /note= "C-protein-specific antibody for immunisation of rabbits"
 FT Region 784..797
 FT /label= P433
 FT /note= "C-protein-specific antibody for immunisation of rabbits"
 FT Region 1007..1020

FT /label= P431
 FT /note= "C-protein-specific antibody for
 FT immunisation of rabbits"

XX US6218142-B1.

XX 17-APR-2001.

XX 05-MAR-1997; 97US-0811583.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASSENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-289830/30.

XX N-PSDB; AAD04370.

XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
 PT polymerase enzymatic activity, useful in modulating gene expression in
 PT plants, humans and animals, as well as in plant cell/tissue cultures or
 PT plant breeding -

XX Claim 1; Column 33-40; 31pp; English.

XX The present sequence is a tomato C-protein having RNA-directed RNA
 CC polymerase (RdRp) activity. This protein is capable of RNA-directed RNA
 CC synthesis, thus using RNA as a template for synthesising complementary
 CC RNA molecules. RdRp nucleic acid is useful for modulating gene expression
 CC in plants, humans and animals. This may lead to various physiological,
 CC developmental and/or morphological changes. Transgenic plants containing
 CC RdRp nucleic acid is especially useful in plant cell or tissue cultures
 CC and in plant breeding. RdRp is useful in gene therapy, particularly for
 CC treating a disease that is caused by the undesirable expression or
 CC overexpression of a gene.

XX Sequence 1114 AA;

Query Match 100.0%; Score 5864; DB 22; Length 1114;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVFGPPYLLSAEYVKSFLKTYGTGTCVCALEYKQSGSRAFAKYQFADNISADK 60

DB 1 MGKTIQVFGPPYLLSAEYVKSFLKTYGTGTCVCALEYKQSGSRAFAKYQFADNISADK 60

QY 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGITLNFQCGISDDKFAVLGSTEVS 120

DB 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGITLNFQCGISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKKFFFLSSGSADYKQLQSYENIWQVYLHRPYGQNAQFLLIQLFGAPRIYKRL 180

DB 121 IQFGIGLKKFFFLSSGSADYKQLQSYENIWQVYLHRPYGQNAQFLLIQLFGAPRIYKRL 180

QY 181 ENSCYSPFKETPDQWVRTTDFPSPWIGLSLSSLCLOFRGRVRLPNPEESFFHYAERENNI 240

DB 181 ENSCYSPFKETPDQWVRTTDFPSPWIGLSLSSLCLOFRGRVRLPNPEESFFHYAERENNI 240

QY 241 TLQTGFTFFYSQKSALPNVQPPGISIPYKILFKISLVLQHGCIQGPALNVYFFRLVDP 300

DB 241 TLQTGFTFFYSQKSALPNVQPPGISIPYKILFKISLVLQHGCIQGPALNVYFFRLVDP 300

QY 301 RRRNVACIEHALEKLYIYKECCYDPVRLWTEQYDGYLKGQPKPSITLDDGLVYVRRV 360

DB 301 RRRNVACIEHALEKLYIYKECCYDPVRLWTEQYDGYLKGQPKPSITLDDGLVYVRRV 360

QY 361 LVTPCKYVFCGPEVNSRVRLNYSIEDIDNLFVSVFDEWEKLYSTDLLPKASTGSGVR 420

DB 361 LVTPCKYVFCGPEVNSRVRLNYSIEDIDNLFVSVFDEWEKLYSTDLLPKASTGSGVR 420

QY 421 TNIIYERILTLRKGFGVIGDKKFFLAFFSSQLRDNVWVFASRPGLTANDIRAWMGDFSQ 480

DB 421 TNIIYERILTLRKGFGVIGDKKFFLAFFSSQLRDNVWVFASRPGLTANDIRAWMGDFSQ 480

DB 421 TNIIYERILTLRKGFGVIGDKKFFLAFFSSQLRDNVWVFASRPGLTANDIRAWMGDFSQ 480

QY 481 IKWAKYAAARLQSGSGSSRETTLSVLRHEIEVIPDVKVHGTSTYVFDGIGKISGDFAHRYA 540

DB 481 IKWAKYAAARLQSGSGSSRETTLSVLRHEIEVIPDVKVHGTSTYVFDGIGKISGDFAHRYA 540

QY 541 SKGGLQYTPSAFQIRYGGYGVVGVDPDSMKLSLRKSKSYESDNKIKLDVLGWSKYQPC 600

DB 541 SKGGLQYTPSAFQIRYGGYGVVGVDPDSMKLSLRKSKSYESDNKIKLDVLGWSKYQPC 600

QY 601 YLNRQLITLLSLGVKDEVLEQKQKEAVDQDAILHDSLKAQEALELMPGENTNLIKAM 660

DB 601 YLNRQLITLLSLGVKDEVLEQKQKEAVDQDAILHDSLKAQEALELMPGENTNLIKAM 660

QY 661 LNCGYKPDAPFLSMMLQTFRASKLLDLTRGRIFTPNGRTMMGCLDESRTLEYGVQVFQ 720

DB 661 LNCGYKPDAPFLSMMLQTFRASKLLDLTRGRIFTPNGRTMMGCLDESRTLEYGVQVFQ 720

QY 721 FTGAGHGFSDDLHPFNNSRSTNSNFILKGNVVAKNPCLHFGDIRVLKAVNVRALHHMV 780

DB 721 FTGAGHGFSDDLHPFNNSRSTNSNFILKGNVVAKNPCLHFGDIRVLKAVNVRALHHMV 780

QY 781 DCVVPQKGRPHNPNCSGSDLDGDIYFYCWQDMIPPRQVOPMEYPPAPSIQLDHDVTI 840

DB 781 DCVVPQKGRPHNPNCSGSDLDGDIYFYCWQDMIPPRQVOPMEYPPAPSIQLDHDVTI 840

QY 841 EEVEEYFTNYIVNDSIGIITANAHVVFADREPDMAMSDPCKLAELFSIAVDFPKTGVPAE 900

DB 841 EEVEEYFTNYIVNDSIGIITANAHVVFADREPDMAMSDPCKLAELFSIAVDFPKTGVPAE 900

QY 901 IPSQLRPKEYPDMKPKDTSYISERVIGKLRKVKDKAPQASSIATFTRDVARSYDAD 960

DB 901 IPSQLRPKEYPDMKPKDTSYISERVIGKLRKVKDKAPQASSIATFTRDVARSYDAD 960

QY 961 MEVDGFEDYIDEAFDYKTEYDNKLGMLNDYIGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020

DB 961 MEVDGFEDYIDEAFDYKTEYDNKLGMLNDYIGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020

QY 1021 SVAVRALRKEARAWFKRRNDIDMLPKASAWHYVTHPTWGCYNQGLKRAHFISFPKCV 1080

DB 1021 SVAVRALRKEARAWFKRRNDIDMLPKASAWHYVTHPTWGCYNQGLKRAHFISFPKCV 1080

QY 1081 YDQLIQIKKDKARNRPNVNLSSLRAQLSHRLVLK 1114

DB 1081 YDQLIQIKKDKARNRPNVNLSSLRAQLSHRLVLK 1114

RESULT 3

AAB28533

ID AAB28533 standard; Protein; 1125 AA.

XX

AC AAB28533;

XX

DT 07-FEB-2001 (first entry)

XX

DE Soybean RNA-directed RNA polymerase, SEQ ID NO: 12.

XX

KW Soybean; plant; RNA-directed RNA polymerase; gene mapping; gene marker;

XX

OS plant virus resistance; plant breeding.

XX

OS Glycine max.

XX

PN WO2000060097-A1.

XX

PD 12-OCT-2000.

XX

XX 06-APR-2000; 2000WO-US09105.

PF

XX 07-APR-1999; 99US-0128094.

PR

XX (DUPO) DU PONT DE NEMOURS & CO E I.

PA (PION-) PIONEER HI-BRED INT INC.

PA


```
Db 622 KEDQSYKSEKILGRYSIQ-----EASNGDVVSQEVCTPNDLPYDIDLEVPGASDFLAS 677
QY 973 AFYKYTEYDNKGLNDYYGIKTEAETLGGIMKASKTFDRRKD--AEALISAVRALRKE 1030
Db 678 AWCKCSYDAQLSALLSQYRTEAEVLVGHITFLVNSKKOGDIKDRLKTAYSALRKE 737
QY 1031 ARAWFK-----RRNDIDMLPKASAWHYHTYHTYGCYNQGL-----KRAH 1072
Db 738 FKSTFESIASDOCEIGDDEKNLLYEM--KASAWYQVTHFK-VWEKSRGILGPDGEEIPA 794
QY 1073 FIFPWCYVDLIQIK--KDKARNRPVLNLSLRQLSHRL 1111
Db 795 SLSEAWIPVDYLARIKLCHGKVRVEGQKPVRLAAYISERI 836

RESULT 8
AAE00898
ID AAE00898 standard; Protein; 218 AA.
AC AAE00898;
XX
XX
DT 04-JUL-2001 (first entry)
DE
DE Tomato RNA-directed RNA polymerase 800 (RdRP800) protein.
XX
XX Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
XX transgenic plant; tissue culture; plant breeding; therapy; C-protein.
XX
XX Lycopersicon esculentum.
XX
XX US6218142-B1.
XX
XX 17-APR-2001.
XX
XX 05-MAR-1997; 97US-0811583.
XX
XX 05-MAR-1997; 97US-0811583.
XX (WASS/) WASSNEGGER M.
XX (RIED/) RIEDEL L.
XX
XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
XX WPI; 2001-289830/30.
XX
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
XX polymerase enzymatic activity, useful in modulating gene expression in
XX plants, humans and animals, as well as in plant cell/tissue cultures or
XX plant breeding
XX
XX Example 3; Column 39-42; 31pp; English.
XX
XX The present sequence is tomato RNA-directed RNA polymerase 800
XX (RdRP800) protein which corresponds to position 700-917 of C-protein.
XX C-protein is capable of RNA-directed RNA synthesis, thus using RNA as a
XX template for synthesizing complementary RNA molecules. RdRP nucleic acid
XX is useful for modulating gene expression in plants, humans and animals.
XX This may lead to various physiological, developmental and/or
XX morphological changes. Transgenic plants containing RdRP nucleic acid is
XX especially useful in plant cell or tissue cultures and in plant breeding.
XX RdRP is useful in gene therapy, particularly for treating a disease that
XX is caused by the undesirable expression or overexpression of a gene.
XX
XX Sequence 218 AA;
XX
XX Query Match 20.2%; Score 1182; DB 22; Length 218;
XX Best Local Similarity 100.0%; Pred. No. 1.6e-97;
XX Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 RTMGCIDESRTLEYGVQVFTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPC 759
Db 1 RTMGCIDESRTLEYGVQVFTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPC 60
```

```
QY 760 LHPGDIRVLKAVNRALHMHVDCVVPQKGRPHNECSGSLDGDIIYFVCDQMIPPR 819
Db 61 LHPGDIRVLKAVNRALHMHVDCVVPQKGRPHNECSGSLDGDIIYFVCDQMIPPR 120
QY 820 QVQPEYPPAPSIQDHDVTTIEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPC 879
Db 121 QVQPEYPPAPSIQDHDVTTIEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPC 180
QY 880 KKLAELFSIAVDFFKTGTGVPAPETPSQIRKPEYDFDMDKP 917
Db 181 KKLAELFSIAVDFFKTGTGVPAPETPSQIRKPEYDFDMDKP 218

RESULT 9
AAB28529
ID AAB28529 standard; Protein; 417 AA.
XX
XX AAB28529;
XX
XX 07-FEB-2001 (first entry)
DT
DE Maize partial RNA-directed RNA polymerase, SEQ ID NO: 4.
XX
XX Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
XX plant virus resistance; plant breeding.
XX
XX Zea mays.
XX
XX WO2000060097-A1.
XX
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09105.
XX
XX 07-APR-1999; 99US-0128094.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX {PION-} PIONEER HI-BRED INT INC.
XX
XX Odeil JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX
XX WPI; 2000-679376/66.
XX N-PSDB; AAC63738.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX for controlling gene expression and providing mechanisms to engineer
XX plant virus resistance
XX
XX Claim 10; Page 34-35; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX isolated from plant cDNA libraries. They are useful as probes for
XX genetically and physically mapping genes, and as markers for
XX traits linked to those genes. They are useful for controlling gene
XX expression and provide mechanisms to engineer plant virus resistance.
XX They are also useful for plant breeding to develop lines with desired
XX phenotypes.
XX
XX Sequence 417 AA;
XX
XX Query Match 17.3%; Score 1013.5; DB 21; Length 417;
XX Best Local Similarity 48.3%; Pred. No. 6.9e-82;
XX Matches 200; Conservative 80; Mismatches 119; Indels 15; Gaps 8;

QY 498 SRTLSVLRHEIEVIPDVKY--HGTSYVPSDGIKTSKSGFAHRVASKGGLQ--YTPSAFC 553
Db 7 SRTQFEVSSYDVEIPDIETDTGTIYFSDGICKITRFAHQVAKLIGLDPAHPPSAFQ 66
QY 554 IRYGGYKGVGVGVPDSDMSKLSLRKSMKESDNIKIDVLGWSKYQPCYLNRLQITLLSTL 613
Db 67 IRYGGYKGVITIDPTSEFFNLSLAPSKMKFKSKTMLNITNWSKQPCYVNRREIISLSTL 126
```

Query Match	11.9%	Score	637.5	DB	21	Length	230
Best Local Similarity	60.1%	Pred. NO.	6.5e-54				
Matches	140	Conservative	30	Mismatches	56	Indels	7
						Gaps	2

QY	386	EDIDNFRV---	SFVDWEWKLSTDL-----	LPK-----	---ASTGSG	418
		:::	:::	:::	:::	
Db	562	EVIOQLMTMGSHSVLGRQWRFFAKDAGYRKPREFOLRAEDPKPIKERVHFEAETGT	621			
		:::	:::	:::	:::	
QY	419	VRTNIYERILSTLRKGFGVIGDKPE--FLATSSSOL-----	RNSVMVFASRPGLPANDI	471		
		:::	:::	:::	:::	
Dh	622	PPDDVEK-----	TRSVVPAPEPYEORTEKVSOMDWLLLOLNNTW----	QHPLXL-----	668	
		:::	:::	:::	:::	

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:44:48 : Search time 21 Seconds
(without alignments)
1560.816 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MKGTIQVFPYLLSAEVVK.....RPVLNLSLRAGLSHRLVLK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5864	100.0	1114	4	US-08-811-583-2
2	1182	20.2	218	4	US-08-811-583-3
3	113.5	1.9	831	4	US-09-269-861A-8
4	112.5	1.9	988	3	US-08-851-843A-69
5	112.5	1.9	988	4	US-08-874-549A-112
6	112.5	1.9	988	4	US-08-854-050-69
7	112.5	1.9	988	4	US-09-430-323-69
8	109	1.9	1215	4	US-09-134-001C-5319
9	105.5	1.8	659	4	US-09-562-737-11
10	105.5	1.8	1194	2	US-08-488-940-1
11	105	1.8	1181	2	US-08-488-940-2
12	104	1.8	1177	3	US-08-754-490-10
13	104	1.8	1177	3	US-08-922-505A-10
14	104	1.8	1177	4	US-09-260-952A-10
15	104	1.8	1177	4	US-09-253-341-10
16	104	1.8	1177	4	US-09-253-331A-10
17	104	1.8	1177	4	US-09-261-040-10
18	103	1.8	1169	4	US-09-255-829-20
19	103	1.8	1177	3	US-08-754-490-28
20	103	1.8	1177	3	US-08-922-505A-28
21	103	1.8	1177	4	US-09-260-952A-28
22	103	1.8	1177	4	US-09-253-341-28
23	103	1.8	1177	4	US-09-253-331A-28
24	103	1.8	1177	4	US-09-261-040-28
25	102.5	1.7	1194	2	US-08-488-940-17
26	102.5	1.7	1626	2	US-08-771-602D-2
27	101.5	1.7	856	4	US-09-134-001C-5438

28	101.5	1.7	1093	4	US-09-315-793-52	Sequence 52, Appl
29	101.5	1.7	1180	3	US-09-224-024-28	Sequence 28, Appl
30	101.5	1.7	1180	5	PCT-US94-07902-28	Sequence 28, Appl
31	101.5	1.7	1626	4	US-09-232-446B-2	Sequence 2, Appl
32	101	1.7	800	2	US-08-488-940-4	Sequence 4, Appl
33	101	1.7	813	2	US-08-488-940-3	Sequence 3, Appl
34	101	1.7	1071	2	US-08-975-527-1	Sequence 1, Appl
35	100.5	1.7	1194	2	US-08-488-940-18	Sequence 18, Appl
36	100	1.7	1177	3	US-08-754-490-25	Sequence 26, Appl
37	100	1.7	1177	3	US-08-922-505A-26	Sequence 26, Appl
38	100	1.7	1177	4	US-09-260-952A-26	Sequence 26, Appl
39	100	1.7	1177	4	US-09-253-341-26	Sequence 26, Appl
40	100	1.7	1177	4	US-09-253-331A-26	Sequence 26, Appl
41	100	1.7	1177	4	US-09-261-040-26	Sequence 26, Appl
42	100	1.7	1661	2	US-08-882-083-2	Sequence 2, Appl
43	100	1.7	1661	2	US-08-558-107-2	Sequence 2, Appl
44	100	1.7	1661	4	US-09-243-539-2	Sequence 2, Appl
45	99.5	1.7	3169	4	US-09-453-702B-257	Sequence 257, App

ALIGNMENTS

RESULT 1

US-08-811-583-2
; Sequence 2, Application US/08811583
; Patent No. 6218142

GENERAL INFORMATION:

APPLICANT: Wassenegeger, Michael
APPLICANT: Riedel, Leonhard
APPLICANT: Schiebel, Winfried
APPLICANT: Sanger, Heinz
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
RNA-DIRECTED RNA POLYMERASE (RDRP)
TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPC-1
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1114 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

Query Match 100.0%; Score 5864; DB 4; Length 1114;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGTIQVFPYLLSAEVVKFLEXTGTGTCALVKSGSRFAKVFQADNISADK 60
|||||

Db 1 MGKTIQVGFPPYLLSAEVKSFLEKTYGTVCALVQKSGSRFAKAVQPADNISADK 60
Qy 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGTLNFCQISDDKFAVLGSTEVS 120
Db 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGTLNFCQISDDKFAVLGSTEVS 120
Qy 121 IQGIGLKKEFFELSSGADYKQLQSYENIOWVVLHRPQGONAQFLLIQLFGAPRIYKRL 180
Db 121 IQGIGLKKEFFELSSGADYKQLQSYENIOWVVLHRPQGONAQFLLIQLFGAPRIYKRL 180
Qy 181 ENSCYSPFKETPDQWRTTDFPSPWIGLSSSLCLOFRRCGRVLPNPFEEFFHYAERENNI 240
Db 181 ENSCYSPFKETPDQWRTTDFPSPWIGLSSSLCLOFRRCGRVLPNPFEEFFHYAERENNI 240
Qy 241 TLQGTFFYFVSOKSALPNVQPEGISIPYKILFKISSLVHOGCIPGALNVYFFRLVDP 300
Db 241 TLQGTFFYFVSOKSALPNVQPEGISIPYKILFKISSLVHOGCIPGALNVYFFRLVDP 300
Qy 301 RRRNVACIEHALEKLYIKKCCYDPVRLWTEQYDGYLKGQPPKSPSITLDDGLVYVRRY 360
Db 301 RRRNVACIEHALEKLYIKKCCYDPVRLWTEQYDGYLKGQPPKSPSITLDDGLVYVRRY 360
Qy 361 LVPCKVYFCGPEVNVNRLVNSEIDNLFVRSFYDEWEKLYSTDLLPKASTGSGVR 420
Db 361 LVPCKVYFCGPEVNVNRLVNSEIDNLFVRSFYDEWEKLYSTDLLPKASTGSGVR 420
Qy 421 TNYERILSTLRKGFVIGDKKFEFLAFSSQLRDNYSVMFASRPGLTANDIRAWMGDFSQ 480
Db 421 TNYERILSTLRKGFVIGDKKFEFLAFSSQLRDNYSVMFASRPGLTANDIRAWMGDFSQ 480
Qy 481 IKNVAKYAARLGQSGFSRSTLVLRHEIEVDPVKVHGTSYVFSQIGKISGDFARHVA 540
Db 481 IKNVAKYAARLGQSGFSRSTLVLRHEIEVDPVKVHGTSYVFSQIGKISGDFARHVA 540
Qy 541 SKCGLQYTPSAFQIRYGYGVGVDPDSSMKLSLRKSMKSYSDNKLKDLVLGWSKYQPC 600
Db 541 SKCGLQYTPSAFQIRYGYGVGVDPDSSMKLSLRKSMKSYSDNKLKDLVLGWSKYQPC 600
Qy 601 YNLQILITLSTLVKQDEVLEQKQKAVDQDLAILHDSLKAQAELEMSGENTNLIKAM 660
Db 601 YNLQILITLSTLVKQDEVLEQKQKAVDQDLAILHDSLKAQAELEMSGENTNLIKAM 660
Qy 661 LNCYKPDABEFLSMLOTPRASKLLDLRTSRIFIPNGRTMMGCLDESRTLEYGVQFVQ 720
Db 661 LNCYKPDABEFLSMLOTPRASKLLDLRTSRIFIPNGRTMMGCLDESRTLEYGVQFVQ 720
Qy 721 FTGAGHGFEFSDLLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLKAVNRALHHV 780
Db 721 FTGAGHGFEFSDLLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLKAVNRALHHV 780
Qy 781 DCVVPQKGRPHNECSGSDLDGDIYFVCDQDMIPPRQVPMYPPAPSIQLDHDTI 840
Db 781 DCVVPQKGRPHNECSGSDLDGDIYFVCDQDMIPPRQVPMYPPAPSIQLDHDTI 840
Qy 841 BEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPCCKLAELFIAVDVFPKTVPAE 900
Db 841 BEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPCCKLAELFIAVDVFPKTVPAE 900
Qy 901 IPSQLRPKEYDQDKDTSYISERVIGKLFVRKVKQKAPQASSIAFTTRDVARSDAD 960
Db 901 IPSQLRPKEYDQDKDTSYISERVIGKLFVRKVKQKAPQASSIAFTTRDVARSDAD 960
Qy 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYGIKTEABILLSGGIMKASKTFDRKDAEI 1020
Db 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYGIKTEABILLSGGIMKASKTFDRKDAEI 1020
Qy 1021 SVAVRALKEARAFKRRNDIDMLPKASANYHYTHPTYGCYNQGLKRAHFTSFPCV 1080
Db 1021 SVAVRALKEARAFKRRNDIDMLPKASANYHYTHPTYGCYNQGLKRAHFTSFPCV 1080
Qy 1081 YDQLIQIKKDKARNRPNVNLSSSLRAQLSHRLVLK 1114
Db 1081 YDQLIQIKKDKARNRPNVNLSSSLRAQLSHRLVLK 1114

RESULT 2
US-08-811-583-3
; Sequence 3, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLIMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811.583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-583-3

Query Match 20.2%; Score 1182; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.3e-112; Indels 0; Gaps 0;
Matches 218; Conservative 0; Mismatches 0;

QY 700 RTMMGCLDESRTLEYGVQFVQFTGAGHGFEFSDLLHPFNNSRSTNSNFIKGNVYVAKNP 759
Db 1 RTMMGCLDESRTLEYGVQFVQFTGAGHGFEFSDLLHPFNNSRSTNSNFIKGNVYVAKNP 60
QY 760 LHPGDIRVLKAVNRALHHVDCVVPQKGRPHNECSGSDLDGDIYFVCDQDMIPPR 819
Db 61 LHPGDIRVLKAVNRALHHVDCVVPQKGRPHNECSGSDLDGDIYFVCDQDMIPPR 120
QY 820 QVQPMYPPAPSIQLDHDTIIEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPC 879
Db 121 QVQPMYPPAPSIQLDHDTIIEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPC 180
QY 880 KKLAEFLSTAVDPKTVGVAEIPSQLRPKEYDPDFMDKP 917
Db 181 KKLAEFLSTAVDPKTVGVAEIPSQLRPKEYDPDFMDKP 218

RESULT 3
US-09-269-861A-8
; Sequence 8, Application US/09269861A
; Patent No. 6468775
; GENERAL INFORMATION:

APPLICANT: Ankenbauer, Waltraud
APPLICANT: Markau, Ursula
APPLICANT: Svetlichny, Vitaly
APPLICANT: Schmitz-Agheguyan, Gudrun
APPLICANT: Reiser, Astrid
APPLICANT: Angerer, Bernhard
APPLICANT: Ebenbichler, Christine
APPLICANT: Laue, Frank
APPLICANT: Bonch-Osmolovskaya, Elizaveta
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE FROM CARBOXYDOTHERMUS HYDROGENOFORMAN
FILE REFERENCE: 4494
CURRENT APPLICATION NUMBER: US/09/269,861A
CURRENT FILING DATE: 1999-11-22
PRIOR APPLICATION NUMBER: PCT/EP97/05391
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: EP/96115873.0
PRIOR FILING DATE: 1996-10-03
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 8
LENGTH: 831
TYPE: PRT
ORGANISM: Carboxydotherrnus hydrogenoformans
US-09-269-861A-8
Query Match 1.9%; Score 113.5; DB 4; Length 831;
Best Local Similarity 19.0%; Pred. No. 0.068;
Matches 158; Conservative 119; Mismatches 298; Indels 257; Gaps 42;
QY 410 LPKASTGSGVTNIYERILSLRKGFGVIGDKKFFELAFSSQLRDNWVWFASRGLTAN 469
DB 20 LPLKLTGEGTGA VYGLTLMFR--VTKDEKPEYLA-----VAFDISKRTFTE 67
QY 470 DIRAMWGFQIKNVAKYAARLGOSFSGSRETSLVLRHEIVPDVKVHGTSYVFSDGIG 529
DB 68 QFTAYKGRKEAPD-----ELVPQFALVREVVLKVN-----VPIYLDG--YEADDIIG 114
QY 530 KISGDA--HRVASKGLOTPSNFQIRYGGYKGV-----GVDPSSMKLSLRKSNK 581
DB 115 HLSRAFAGOGHEV-----VIYADRMQLQVDEKTVVLYLTKGI--TELVKMDLAAILN 167
QY 582 Y-----ESDNKIDVLGWSKYQPCYLNRLQITLLSTLGVRDELEOKOK 625
DB 168 YGLKPKQLVDYKGLMGDFSDNI-----FGVPGIGKGTALDLKITYGSVEVLARK-- 217
QY 626 EAVDQDAILHDSLKAQBALEIMSPGENTNLIKAMLC-----GYKPDREPFLSM 675
DB 218 ---DELKPKLREKLAHENLAKISKQLATILREIPLEISLEDLVKKEPNYEYEVAKLFLH 274
QY 676 MLQTRASKLLDLRTSRIFTPNGRTMGCIDESRTLE-YGVQVFOFTGAGHGFSDDLH 734
DB 275 EFKSP-----LKEIEPKIKKEVQEGKDLV-----QVETVETEGQIAVFSDFYVDDGKTK 326
QY 735 PFNNRSSTNSNFIKGNVVVAKNP-----CLHPG-----DIRVLKAV-----N 772
DB 327 FYSLDRLNEIEIFRNKIIITDARKGIYHVCKLEKLTPEVCFDARIAAYVLNPAQNPG 386
QY 773 VVALHWMDCVFPQKGRPHNECSGSDLDGDIYF-----VCDQDMIPROVQPMEX 826
DB 387 LKGLYLKLDLPY-----EDVSLNIRGLFYLKEMMRKPIFEQO--ERLFYEIEL 434
QY 827 PPAP-----SIQLDHDVTI-----EEVEEYFTNYIVNDSLGIITANARVVFADREP 871
DB 435 PLTPVLAQMEHTGIQVDREALKEMSLGEQIEE-----LIREIYVLAGE--- 479
QY 872 DNAMSDPKKLAELSIAVDFPKTGV-----AEIPSLRP----- 907
DB 480 EFNLSF-RQLGVIL-----FEKGLPLVTKTKTGYSDAEVLBELLPFHEIGIKILNY 533
QY 908 -----KEYPDMFKDPKTSYISERV-----IGKLFKRVKDKAPQASSIATFTRD 951
DB 534 RQLMKLKSYITDLK-----PLINERTGKLHTFTNQGTILTORLASSPENLQNI-PIRL 586

QY 952 VARRSYDADMEVDGFEDYIDEAFDYKTEYDNKLNLMYDYGKTEABEILSGGIMKASKTF 1011
DB 587 LGRKLRKMEIPSPGY-DIIVSA-----DYSQIELRLLAHF--SEEPKLEA--YQKGEDI 636
QY 1012 DRKDAEAI5AVRALRKEARAFKERN-----DIDDM-----LPKASAWHTVHT 1059
DB 637 HRKTASEVFGVLEEVTPEMRAHAKSVNFGIVGSDFGDLGDLKIPREVAKYI----- 591
QY 1060 YMGYNQGLKRAHFISFPWC--VYDOLIQIKDK-----ARNRPVLNLS 1102
DB 692 -----KNYFANYPKVREVLDLVTAREKGVYVTLFGRRRYPPELSS 733
RESULT 4
US-08-851-843A-69
Sequence 69, Application US/08851843A
Patent No. 6093809
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6093809el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-851-843A-69
Query Match 1.9%; Score 112.5; DB 3; Length 988;
Best Local Similarity 18.2%; Pred. No. 0.12;
Matches 128; Conservative 109; Mismatches 258; Indels 207; Gaps 30;

```
QY 134 LSSGSADYKQLQSYENIWQVVLHRYPGQNAQFLLIQLFGAPRIYKRLNSCYSFFKRPD 193
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 130 LVSTFPNYLISLESKNWQ-----LLLEIGSDAMHYLLSKG--SIFEALPN 174
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 194 DQWVRTTDFP-----PSWIGLSSSLCLOQPRGV 221
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 175 DNYLQISGIPLFKNVFEETVSKKRRTIETITQNKSAKKEYSWNSISISRSFIYRS 234
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 222 RLPNFESFPHYAERENNITLQGTFFVFSQKALVPNVQPPGSIPIPKILFKISSLV- 380
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 235 YKFKQDLYFNHLSICDNTVHMWLFPPQFGLNAPQ-----VKQLHKVPLVS 286
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 281 QHGCIPGALNVYFRLVDPDRRNVACIEHALEKLYIKECCYDVPVRLTEQDGYLKG 340
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 287 QSTVVPKRLKVY--PLIEQTAKRLHRI--SLSKVY-----NHYPYIDTH 328
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 341 QPKPSPSITLDDGLV--YVRRVLVTPCKVFCGPEVNVSN--RVLRNYSIEDIDNFLRVSFV 397
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 329 DDEKILSYSLKPNQVFAFLRSILV--RVF--PKLWGNQRIEILKLDLFTFLSLRY 382
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 398 DE-----EWEKLYSTDLKPASTGSGVTRNIYERILSTLRKGVIGDKKF 442
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 383 ESFSLHYLMSNIKISEIENLVGKRSNAKMLCLSDFEKRRQIFAEFIYWLNSFIPILOS 442
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 443 EFLAFSSSLDRNSV-----WMFASPPGLTANDIRAW----- 474
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 443 FFYITESDIRNRVYPRKDIWKLLCRPFTSMKMEAFKINENNVNMDQKTLTPPAVI 502
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 475 -----MGDFSQIKNAK--YAARLGQS---FGSSRETL-----SVLRH-----EIEVIP-DVK 516
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 503 RLLPKKNTFRILTNRFRFKIMGSKNKLMLVSTNOTLRPVASILKHLINDESSGIPFNLE 562
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 517 VHGTSYVFSGIGISGDFAHRYASKGLOQYTPSAFOIRYGVYGVVGVDPDSMKLSLR 576
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 563 VYMKLLTFKDDLLK-----HRM-----FGKKYFVRIDIKCY----- 595
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 577 KMSKYPESDNKLD-----VLGWSKYQPCYLNRLQITLSTLGVKDEVLEQKQKEAVDQLD 632
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 596 -----DRIKQDLMFRIVKKLKDPEFVIRKYATTHAT---SDRATKNFVSEAFSYFD 644
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 633 AILHDSLKAQALPMSFGENTNLKAMLCNGYKPDAPLMSMLOFRASKLLDL---R 689
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 645 MV-----PFEKVQLLSMKYSITLFDVDFVDFVWTKSSSEIF--KMKELHSGHIVKIGNSQ 697
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 690 TRSRIFIPNGRT-----MMGCLDE--SRTLEYQGVFVQ 720
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 698 YLQKVGIPQGSILSSFLCHFYMEDILDEYLSFTKKKGSVLLR 739
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

RESULT 5
US-08-974-549A-112
; Sequence 112, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Greg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 23-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-0026100S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-974-549A-112

Query Match 1.9%; Score 112.5; DB 4; Length 988;
Best Local Similarity 18.2%; Pred. No. 0.12;
Matches 128; Conservative 109; Mismatches 258; Indels 207; Gaps 30;

QY 134 LSSGSADYKQLQSYENIWQVVLHRYPGQNAQFLLIQLFGAPRIYKRLNSCYSFFKRPD 193
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 130 LVSTFPNYLISLESKNWQ-----LLLEIGSDAMHYLLSKG--SIFEALPN 174
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 194 DQWVRTTDFP-----PSWIGLSSSLCLOQPRGV 221
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 175 DNYLQISGIPLFKNVFEETVSKKRRTIETITQNKSAKKEYSWNSISISRSFIYRS 234
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 222 RLPNFESFPHYAERENNITLQGTFFVFSQKALVPNVQPPGSIPIPKILFKISSLV- 380
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 235 YKFKQDLYFNHLSICDNTVHMWLFPPQFGLNAPQ-----VKQLHKVPLVS 286
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 281 QHGCIPGALNVYFRLVDPDRRNVACIEHALEKLYIKECCYDVPVRLTEQDGYLKG 340
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 287 QSTVVPKRLKVY--PLIEQTAKRLHRI--SLSKVY-----NHYPYIDTH 328
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 341 QPKPSPSITLDDGLV--YVRRVLVTPCKVFCGPEVNVSN--RVLRNYSIEDIDNFLRVSFV 397
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
Db 329 DDEKILSYSLKPNQVFAFLRSILV---RVF---PKLIWGNQRIFEIILKOLETEFLKLSRY 382
QY 398 DE-----EWEKLYSTDLLPKASTGSGVRTNIYERILSTLRKGFVIGDKKF 442
Db 393 ESFSLHYLMSNIKISEIWLVLKRSNAKMCGLSDFEKKQIAPFEIYWLNSFIPILOS 442
QY 443 EFLAFSSQLRDSNV-----WMFASRPGLTANDIRAW----- 474
Db 443 FFYITESSDLNRRTVYFRKDIWKLLCRPFITSMKMEAFKINENNVRMDTKTTLPPAVI 502
QY 475 -----MGDFSQIKNVAK-YAARLQGS-----FGSSRETL-----SVLRH-----EIEVIP-DVK 516
Db 503 RLLPKKNTFRITNLKRFLKMGSKMKMLVSTNQTLRPVASILKHLINESGIPFNLE 562
QY 517 VHGTYSYVSDGIGKISGDFAHRYVASKGLQYTPSAFYRYGYGVGVGVDPSMKLSLR 576
Db 563 VYMKLLTFKKDLLK-----HRM-----FGKKYFVRIDIKSCY----- 595
QY 577 KSMKSYESDNKILD-----VLGWSKYQPCYLNRLQITLLSTLGVKDEVLEQKQKEAVDOLD 632
Db 596 -----DRIKQDLMPRIVKKLKDPEFVIRKYATIHAT---SDRATKNFVSEAFSYFD 644
QY 633 AILHDSLKAQEALELMSPGENTNLIKAMLCNGYKPDAPFELSMMLQTFRASKLLDL---R 689
Db 645 MV-----PFEKVYQLLSMKTSDTLFVDFVDYWTKSSSEIF--KMLKEHLSGHIVKIGNSQ 697
QY 690 TRSRIFIPNGRT-----MMGCLDE--SRILEYGQVFVQ 720
Db 698 YLQKVGIPQGSILSSFLCHFYMEDLIDELYSFTKKKGSVLLR 739

RESULT 6
US-08-854-050-69
; Sequence 69, Application us/08854050
; Patent No. 6261836
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6261836el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
```

```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-0029300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 988 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-854-050-69
```

```
Query Match 1.9%; Score 112.5; DB 4; Length 988;
Best Local Similarity 18.2%; Pred. No. 0.12;
Matches 138; Conservative 109; Mismatches 258; Indels 207; Gaps 30;

QY 134 LSSGSADYKILQSYENINQVVLHRYPCQNAQFLLIQLFAGAPRIYKRLNESCYSFFKETPD 193
Db 130 LVSTFPNYLISLESKNWQ-----LLEIGSDAMHYLLSKG--SIFPALPN 174
QY 194 DQWVRTTDFP-----PSWIGLSSSLCLOFRRGV 221
Db 175 DNYLQISGIPLFKNVPEETVSKKRTIETSIQNSARKEVSWNSISISRSFIFRSS 234
QY 222 RLNFESFPHYAERENNILOTGFTFFVQSXSALVNPVQPPGISIPYKILEKISLV- 280
Db 235 YKFKQDLYENLHSDICDRNTVHMWLOTFFRQGLNAFO-----VKLHKVITPLVS 286
QY 281 QHGCIQGPALNVYFFRLVDPRRRNACIEHALEKLYIKECYDVPWMLTEQDGYLKR 340
Db 287 QSTVVPKRLKVV--PLIEQTAKRLHRI--SLSKVY-----NHICYPIDTH 328
QY 341 QPPKSPSITLDDGLV--YVRVLVTPCKYFCGPEVNVSN-RVLKNYSEIDNFLRVSVF 397
Db 329 DDEKILSYSLKPNQVFAFLRSILV---RVF---PKLIWGNQRIFEIILKOLETEFLKLSRY 382
QY 398 DE-----EWEKLYSTDLLPKASTGSGVRTNIYERILSTLRKGFVIGDKKF 442
Db 383 ESFSLHYLMSNIKISEIWLVLKRSNAKMCGLSDFEKKQIAPFEIYWLNSFIPILOS 442
QY 443 EFLAFSSQLRDSNV-----WMFASRPGLTANDIRAW----- 474
Db 443 FFYITESSDLNRRTVYFRKDIWKLLCRPFITSMKMEAFKINENNVRMDTKTTLPPAVI 502
QY 475 -----MGDFSQIKNVAK-YAARLQGS-----FGSSRETL-----SVLRH-----EIEVIP-DVK 516
Db 503 RLLPKKNTFRITNLKRFLKMGSKMKMLVSTNQTLRPVASILKHLINESGIPFNLE 562
QY 517 VHGTYSYVSDGIGKISGDFAHRYVASKGLQYTPSAFYRYGYGVGVGVDPSMKLSLR 576
Db 563 VYMKLLTFKKDLLK-----HRM-----FGKKYFVRIDIKSCY----- 595
QY 577 KSMKSYESDNKILD-----VLGWSKYQPCYLNRLQITLLSTLGVKDEVLEQKQKEAVDOLD 632
Db 596 -----DRIKQDLMPRIVKKLKDPEFVIRKYATIHAT---SDRATKNFVSEAFSYFD 644
QY 633 AILHDSLKAQEALELMSPGENTNLIKAMLCNGYKPDAPFELSMMLQTFRASKLLDL---R 689
Db 645 MV-----PFEKVYQLLSMKTSDTLFVDFVDYWTKSSSEIF--KMLKEHLSGHIVKIGNSQ 697
QY 690 TRSRIFIPNGRT-----MMGCLDE--SRILEYGQVFVQ 720
Db 698 YLQKVGIPQGSILSSFLCHFYMEDLIDELYSFTKKKGSVLLR 739
```

RESULT 7
US-09-430-323-69
; Sequence 69, Application US/09430323
; Patent No. 6309867
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin
; Andrews, William H.
; TITLE OF INVENTION: No. 6309867el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/430,323
; FILING DATE: 29-Oct-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 988 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 69:
US-09-430-323-69

Query Match 1.9%; Score 112.5; DB 4; Length 988;
Best Local Similarity 18.2%; Pred. No. 0.12;
Matches 128; Conservative 109; Mismatches 258; Indels 207; Gaps 30;
QY 134 LSSGSADYKQLQSYENINQVWLHRRPYGONAFLIQLFGAPRIYKRLNSCYSFETPD 193
DB 130 LVSTFPNYLISLESKNQ-----LLEIIGSDAMHYLSKG--SIFEALPN 174
QY 194 DQWRTTDFP-----PSWIGLSSSLCLQFRGV 221
DB 175 DNYLQISGIPKNNVFEYTSKKKRIETSIQNKSRKEVSNWSISISFIFESS 234
QY 222 RLNPNEESFFHYAERENITLQGTFFVSQKSAIPVWQPEGISIPYKILFISSLV- 280
DB 235 YKKFKQDLYFNLSICDNTVMWLQWTFPQFGLINAFQ-----VKQLHKVPLVS 286

QY 281 QHGCIPGALNVYFFRLVDPRRNRVACIEHALEKLYIKECCYDVRMLTEYDYDYLKGR 340
DB 287 QSTVVPKRLKVV--PLIEQTAKRLHRI--SLSKV-----NHYCPYIDTH 328
QY 341 QPPKSPSITLDDGLV--YVRVLVTPCKYFCGPEVNVSN-RVLNYSIEDIDNLRVSEV 397
DB 329 DDEKILSYSLKPNQVFAFLRSILV---RVF---PKLWGNQRIFEILLKDLTFKLRSY 382
QY 398 DE-----EWEKLYSTDLLPKASTGSGVTRTIYERILSTLRKGFVIGDKKF 442
DB 383 ESFSLHLYMSNIKISIEIWLVLGKRSNAKCLSDPEKRKQIFAETIYWLNSFIPIQS 442
QY 443 EPLAFSSQLRNSV-----WVFASRPGLTANDIPAW----- 474
DB 443 FFYITESSDLRNRVTYFRKDINKLRCRPFITSMKMEAFKINENNVMVMDTQTTTPPAVI 502
QY 475 -----MGDFSQIKNAK-YAARLQGS---FGSSRETEL---SVLRH---EIEVIP-DVK 516
DB 503 RLLPKNTFRLLTNLRKFLIKMGSNNKMLVSTNQTLRPVASILKHLINEESSGIPFNLE 562
QY 517 VHGTSYVSDGIGKISGDFAHVASKCGLOYTPSAQIRYGGYGVGVGVDPSSMKLSLR 576
DB 563 VYMKLLTFKKDLLK-----HRM-----FGRKKYFVRIDIKSY----- 595
QY 577 KMSKYESDNILK-----VLGWSKYQPCVNLRLITLLSTLGVKDEVLEQKQEAVDQLD 632
DB 596 -----DRIKQDLMFRIVKVKLKDPEVIRKYATIHAT---SDRATKNFYSEAFSYD 644
QY 633 AILHDSLKAQEALELMSGENTNILKAMLCNGYKPDAPFPLSMLOTFRASKLLDL---R 689
DB 645 MV-----PFEKVYQLLSMKTSDTLFVDVYWTKSSEIF--KMLKEHLSGHIVIGNSQ 697
QY 690 TRSRIPNGRT-----MMGCLDE--SRTLEYGVQVFO 720
DB 698 YLOKVGIPQGSILSFLCHFYMEDLIDEYLSFTKKKGSVLLR 739
RESULT 8
US-09-134-001C-5319
; Sequence 5319, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5319
; LENGTH: 1215
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5319
Query Match 1.9%; Score 109; DB 4; Length 1215;
Best Local Similarity 19.1%; Pred. No. 0.4;
Matches 117; Conservative 85; Mismatches 203; Indels 208; Gaps 31;
QY 539 VASKCG--LQYTPSAQ--TRYGGYGVGVGVDPDSSMKLSLRKSMKYESDNIKLDVLGW 594
DB 118 IPSRMGILLDMSPRALEVIFYASY---VVVDGPT---GLEKKTLLSEAE-----F 163
QY 595 SKYQPCVNLRLITLLSTLGVKD---EV-LEQKQEAVDQDAILHDSL-KAQEALELMS 649
DB 164 REYDKYPN-QFVAKWGAEGIKLLEIDLEBELKELRDELESATGQRLTRAIKRLVEVE 222
QY 650 PGENTNILKAMLCNGYKPDAPFPLSMMLQ-----TFRASKLLDLRTR----- 691

Db 223 SFRNGNPNMILDLVLP1PPIRPMVQDGGFATSDLDLYRRVIRNRRNLKRLLDL 282
QY 692 --SRIFPNGRMTMGCLDES--RTLEYGQVFOFTGAGHGEFSDDLHPFNNSRSTNSFI 747
Db 283 GAPILIVQNEKRM--LOEAVDALIDNGRGRPVTPGPNPLKSLSHMLKAGKGRFRQNL 339
QY 748 L-----KGNVYVAKNP-----CLHFGDIR-----VLK-----AVNVR----- 774
Db 340 LGRVDYSGRSVIAVGPGLKMYQGLPKEMALELFPKPVFMKELVQREIATNIKNNAKSIE 399
QY 775 -----ALJHM-----VDCVFPQKGRPH--NECSGSDLD 803
Db 400 RMDDEVWDLVDVTEHPVLLNRAPTLHRIGIQAFETLVEGRAIRLHLPLVTTAYNADF 459
QY 804 GDIYFV-----CWDQMIPPRQVQMPYPPAPSIQLDHVDVTEEEVEEYF 847
Db 460 GDQMAVHPLSKAQAEARMMLAAQILNPKDKGKPVVTPSQDMVLGNLYTLTKERDAVN 519
QY 848 TNYIVNDS-----LGIANA--HVVFADREPDMAMSDCKKLAEFSI 888
Db 520 TGAIFNDTNEVLKAYGVHHLTRIGVHANSFNNFTTDEQSKILATSVGKI--IFNE 577
QY 889 AVDFPKTGVAEIPSQLRPKPEYDFMDPKDKTYSISERVIGKLFKRVKDKAPQASSIATF 948
Db 578 II-----PDSFAYINEPSQANLERTTPDK----- 601
QY 949 TRDVARSYDAD--MEVDGFEDYIDAFDYKTEYDNK-LGNLMDYYIGIKTEAELISG-G 1003
Db 602 -----YFVDP7QLGEGGLKEYFDNA-ELIEPFNKKFLGNII-----AEVFNRF 644
QY 1004 IMKASKTFDRKD 1016
Db 645 ITDISMLDRMKD 657

RESULT 9

US-09-562-737-11
; Sequence 11, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSN0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-11

Query Match 1.8%; Score 105.5; DB 4; Length 659;
Best Local Similarity 20.5%; Pred. No. 0.29;
Matches 67; Conservative 46; Mismatches 119; Indels 95; Gaps 16;
QY 791 RHPNECSGSLDGGIYFVCDQMIPPRQVQPM-----EYPPAPSIQLDHVDVTE 841
Db 320 RHP-----SISEDEGTDC-----LSSPERAEPGGGWRGSLGEPPPPRASSDTSAL 370
QY 842 EYEEYFTNYVNDSLGIIATANHV-----FADREPDMAMSDPKKLAEFSIADVFP 893
Db 371 SYDSKYVLIVDEH-----AQLELWSLRPCFGDYSDESATVYNCASASSPYESAI--- 423
QY 894 KTGVAEIPSQLRPKPEYDFMDPKDKTYSISERVIGKLFKRVKDKAPQASSIATF---- 949
Db 424 --GEEYEAQPRPTCLSEDSTDPDVFSSKKEL-NVFMGSRGRSSSAESFGLFSCVIN 480

QY 950 -----ROVARRSYDADMEVDG-----FEDYIDEAFDYKTEYDNKLGMLMDY 990
Db 481 GEEHEQTHAIRFRFVPRHEDELEVEDDPLLVLELQAEADYWEYAYNMRT---GARGVFPAY 537
QY 991 YGKLT--EAEILSGGIMKASKTFDRKDAEATSVAVRALRKEARAWFKRRNDIDDMPL-- 1046
Db 538 YALEVTKPEHMAA-----LAKNSCVLEISVRGVKIGVKA-----DDALEAK 579
QY 1047 --KASAWYHVT-----YHP---TYWG 1062
Db 580 GNKCSHFFOLKNISFCGYHPKNNKYFG 606

RESULT 10

US-08-488-940-1
; Sequence 1, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1194 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-940-1

Query Match 1.8%; Score 105.5; DB 2; Length 1194;
Best Local Similarity 19.9%; Pred. No. 0.89;
Matches 176; Conservative 114; Mismatches 360; Indels 235; Gaps 43;
QY 218 RRGVRLPNFEE-SFFHYAERENNTLQGTFFF-----VSQKSALVPNVQPPPEG-ISIP 269
Db 326 QKGEIMPNIPOMSAFWAVRTAVTAVINAASRQTVDKALQKQATNSSSVYGRGSIAGP 385
QY 270 YKIL---FKISSLVQH--GCIPG--PALNVYFRL-----VDPRRRNVA 306
Db 386 EWLDRPVSNNSQLVSVAGTVEGTNQDISLKFFEDLITSRPAHGKTEGGLSPKSKPFA 445
QY 307 ---CIBHALEKLYIKECCYDPRVWLTEQY-----DGYLKGRQPPKSPSITLDDGLV 355
Db 446 TDSGAMSHKLEKA-----DILKAIQEQOLIANHNSDDYFEVIDFASDATITDRNGV 497
QY 356 YVR-----RVLYTPCKVYFCGPEVNV--SNRVLRYNSEDIDNPLRVSVFDEWEKLYST 407
Db 498 YFADKDGSVTLPTQPVQEFLLSGHVRVRYRKEKPIQNAQKSDVDEYTVQF-----T 547

```
QY 408 DLLPKASTGSGVRNIYERILSTRKGFVIGDKKFEFLAFSSQLRDNSSVWFASRPGLT 467
Db 548 PLNPDDEFRPLKDT---KLLTLAIGDTITSQ--ELLAQAQSIILNKN-----HPGYT 595
QY 468 A-----NDIRAWMGDFSQ-----IKNVAKAARLGQSGSRETSLVLRHEIEV 511
Db 596 IYERDSSIVTHDNDIFRILPMQDEFTYRVKN-REQAIRINKKSGLNEE-----643
QY 512 IPDVKHGTSYVDSFGIKIGSGFAHRVASKGGLQTP-----SAFQIRYGGYKGVGV 565
Db 644 -----INNTDLISE-----KYVVKKEKPYDPDRSHLKLFTIKY-----V 680
QY 566 DPSSMKLSLRKSKSYESNIKLDVLGWSKYQPCYLNRQLITLLSTGLVKDEVLEQKQK 625
Db 681 DVDTN-----ELLSQTLASERNLDFR--DLYDPRDKAKLLYNLDAFGIMDYTLTGKVE 735
QY 626 EAVDQLDAILHDSL-KAQEALELMSPGENTNILKMLNCGYKPDABPFLSMMLQTFRASK 684
Db 736 DNHDDTNRIITVYMGKRPE-----GENASYHLAYDKDRYTEEREVYSYLRYT--GTP 786
QY 685 LLDL---RTSRIFIPNGRTMGLDESRTLEYQGVFVQFTGAGHGEFSDDLHPFNNSRS 741
Db 787 IPDPNPKNNSQLVSVAGTVEG-TNQDISLKFEIDLTSRPAHGKTEQGLSPKSKPFA 845
QY 742 TNSNFILKGNVYVAKNPCLHPGDIRVLKAV-----NVRALHEMVDQVFPQKGRPHPN 795
Db 846 TDSG-----AMSHKLEKADL--LKAIQEQLIANVHSNDYFVIDF-----AS 886
QY 796 ECGSGLDGIYFCWDDMIPPRQVOPMEYPPAPSIQLDHDVYIEVEEYFNYIVNDS 855
Db 887 DATITDRNGKVFYFADKDGSVTLPTQ-----PVQEPFLSGHVRV-----RYKEKPIQNOA 935
QY 856 LGLIANAHVVFADREPDMAMSDPCK--KLAELFSIAVDFPKTGVPAPISQLRKPEYDF 913
Db 936 KSVDEVTVQFTPLNPDDFRPGLKDTKLLKTLAIGDTITSQELLAQAQSIILN-KNHPGY 984
QY 914 MDXPKDTSYISERYIGLFRKVKDKAPQASSIATFTREVARRSYDADMEVDGFDYIDEA 973
Db 995 T-----IYER-----DSSIVTHDNDIFRILPMQDE-----1020
QY 974 FDYKTEYDNKLGMLDYYGIKTE---AEILSGG---IMKASKTFD 1012
Db 1021 FTYRVKREQAYRINKKSGLNEEINNTDLISEKYIYVLKKEKPYD 1065

RESULT 11
US-08-488-940-2
; Sequence 2, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMEUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
```

REFERENCE/DOCKET NUMBER: 05433/009001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1181 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-488-940-2

Query Match 1.8%; Score 105; DB 2; Length 1181;

Best Local Similarity 19.9%; Pred.No.0.98; Indels 228; Gaps 42;

Matches 174; Conservative 114; Mismatches 359; Indels 228; Gaps 42;

```
QY 218 RRGVRLNPTEE-SFFHYAERENNTITQTG-FTFFVSQKSALVPNVQPPEGISIPYKILFK 275
Db 326 QKGEIMPNIPOMSAFWYAVRTAVINAAAGSQFVDEALQDAQTNSSSVPGSGIEGR--N 382
QY 276 ISSLVQH--GCTPG--PALNVTFFRL-----VDPRRNVA---CIEHAL 312
Db 383 NSQLVSVAGTVEGTNQDISLKFEIDLTSRPAHGKTEQGLSPKSKPFAIDSGAMSHKL 442
QY 313 EKLYIKKECCYDPVRLWLTEQY-----DGYLKGRQPKPSITILLDGLVYVR-----R 359
Db 443 EKA-----DLLKAIQEQLIANVHSNDYFVIDFASDAITIDRNGKVIYFADKDGSVT 494
QY 360 VLVTCPKVYFCGPEVNV--SNRVLRNYSIEDINFLRVSPVDEWEKLYSTDLLPKASTGS 417
Db 495 LPTQPVQEPFLSGHVRVRYKKEPIQNAKSDVVEYTVQF-----TLPNPDDFRP 544
QY 418 GYRTNIYERILSTRKGFVIGDKKFEFLAFSSQLRDNSSVWFASRPGLTA-----468
Db 545 GKLDT---KLLTLAIGDTITSQ--ELLAQAQSIILNKN-----HPGYTIYERDSSIYV 592
QY 469 --NDIRAWMGDFSQ-----IKNVAKAARLGQSGSRETSLVLRHEIEVDPVKVHGTS 521
Db 593 HNDIFRTILPMQDEFTYRVKN-REQAIRINKKSGLNEE-----INNTD 635
QY 522 YFSDGIGIGSGFAHRVASKGGLQTP-----SAFQIRYGGYKGVGVDPDSSMKLSL 575
Db 636 LISE-----KYIYVLKKEKPYDPDRSHLKLFTIKY-----VDVDIN--EL 674
QY 576 RKSMKSYESDNTKLDVLGWSKYQPCYLNRQLITLLSTGLVKDEVLEQKQEAVDQLDAL 635
Db 675 LKSEQLLTASERNLDFR--DLYDPRDKAKLLYNLDAFGIMDYTLTGKVEDHDDTNRII 732
QY 636 HDLSL-KAQEALELMSPGENTNILKMLNCGYKPDABPFLSMMLQTFRASKLLDL---RTR 691
Db 733 TVYMGKRPE-----GENASYHLAYDKDRYTEEREVYSYLRYT--GTFIPDNPNDKNN 783
QY 692 SRIFIPNGRTMGLDESRTLEYQGVFVQFTGAGHGEFSDDLHPFNNSRSTNSFILKN 751
Db 784 SQLVSVAGTVEG-TNQDISLKFEIDLTSRPAHGKTEQGLSPKSKPFAIDSG-----836
QY 752 VYVAKNPCLHPGDIRVLKAV-----NVRALHEMVDQVFPQKGRPHNEGSGSDLDG 805
Db 837 ---AMSHKLEKADL--LKAIQEQLIANVHSNDYFVIDF-----ASDAITIDRNGK 883
QY 806 IYFVCWDDQMIPPRQVOPMEYPPAPSIQLDHDVYIEVEEYFNYIVNDSLGIANAHVY 865
Db 884 VYFADKDGSVTLPTQ-----PVQEPFLSGHVRV---RYKEKPIQNAKSDVVEYTVQ 932
QY 866 FADREFDMAMSDPCK--KLAELFSIAVDFPKTGVPAPISQLRKPEYPPFMDKPKDTSYI 923
Db 933 FTPLNPDDFRPGLKDTKLLKTLAIGDTITSQELLAQAQSIILN-KNHPGYT-----I 983
QY 924 SERVIGKLFKRVKDKAPQASSIATFTREVARRSYDADMEVDGFDYIDEAFDYKTEYDNK 983
Db 984 YER-----DSSIVTHDNDIFRILPMQDE-----FTYRVKREQ 1017
```

QY 984 LGNLMXYGIKTE---ABILSG---IMKASKTD 1012
Db 1018 AYRINKKSGLEINNTDLISEKYYVLKKEKPYD 1052

RESULT 12

US-08-754-490-10
; Sequence 10, Application US/08754490
; Patent No. 6017534
; GENERAL INFORMATION:
; APPLICANT: Malvar, Thomas
; APPLICANT: Gilmer, Amy Jelen
; TITLE OF INVENTION: HYBRID BACILLUS THURINGIENSIS
; TITLE OF INVENTION: DELTA-ENDOTOXINS WITH NOVEL BROAD SPECTRUM
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/754,490
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: MOBT:009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-754-490-10

Query Match 1.8%; Score 104; DB 3; Length 1177;
Best Local Similarity 18.2%; Pred. No. 1.2; Mismatches 340; Gaps 45;
Matches 168; Conservative 112;
QY 102 FGCQISDDKFAVLGSTEVS-----IQFGIGLKKFFFLSSGSADYKLOLSYENIWQVYLH 156
Db 13 YNC-LSNPEVEVLGGERIETGTPIDISLSLQFLSEVPVCGAGFVLGL-VDLIWI--- 67
QY 157 RPYG--QNAQFL-----IQFGAPRIYKRL--NSCYSPFKETPDQWVTTDEP 203
Db 68 --FGPSQWDAFLVQIEQLINQRIEFAFNQAIISRLGSLNLYQIYAES-----FREWEAD 120
QY 204 PSWIGLSSSLCQFRGRVGLPNFESEFFHYAERENNTQTGTFEFVSOKSALVFNVPQP 263
Db 121 PNPALREWRQF-----NDMNSALTTPLEA-----VQNYQVP 156
QY 264 -----EGISIPYKILFKISLVHGCIPGPNL----- 291
Db 157 LLSVYVQAANLHLSVLDRVSFGQWGFDAATINSRYNDLTRLIGNYTDYAVRWYNTGLE 216
QY 292 -----VYFFRLVDPRRNVACIEHALEKLYIKKECCVD 324
Db 217 RVWGPSPDRWVRVNPQFRELTTVLVDIVALEPNYDSDRYPIRTVSQLTREIY-----TN 270
QY 325 PVRWLTEQDGYLKG-----RQPPKSP-----SITL-----DGLVY--VRRVLVTPCK 366

Db 271 PV---LENFDGSPRGSAGQIERSPHLMIDILNSITIIDAHRYYYNSGHHIMASP--- 325
QY 367 VYFCGPEVNVYSNRYLNRNYSEDIDNFLRVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420
Db 326 VGFSGPEFTE-----PLYGTMGNAAPQQRIVAQLGGV- 358
QY 421 TNIYERILSTL-RKGFVIGDKKEFLAFSSOLR--DNSVWMEFASRPGLTA----- 468
Db 359 ---YRTLSSTLYRRFPNIG-----INNQLSLVDGTEFAYGTSNLSNLSAYVRKSGTV 407
QY 469 -----NDIRAWMGDFSQIKNVAAYARLGOSFGSSRRETLVSLR----- 506
Db 408 DSLDEIPQNNVPPRQGFSHRLSHVSMFRS-----GFSNSSVSIIRAPMFSWTHRSAT 461
QY 507 -----HELEVIPDKVH-----GTSYVFSDDGI--GKI-----SGDFAHRVASKCGLOYT 548
Db 462 PTNIDIPERITQPLVKAHTLQSGTTVVVRGPGFTGGDILRRTSGGPFAYTIVININGOLPQ 521
QY 549 PSAFQIRYGG-----YKGVGV-----DPDSSMKLSLRKMSKY----- 582
Db 522 RYRARIYASTNLRIYVTVAGERIFAGQFNKMTMDTGDPLTFQSFVSATINTAFTEPMSQ 581
QY 583 ESDNIKLDVLGWSKYQPCYLN-----OLITLLSTLGVKDE 618
Db 582 SSFTVGADT--FSSGNEVYIDRFELIPVTATFEAYDLERAQKAVNALFTSINQIGIKTD 639
QY 619 VLEQKQKAEVDQDAI---LHDSLKAQEALELMSPGENTNILKAMLNCGYKYPDAEPFLSM 675
Db 640 VTDVH---IDQVSNLYDCLSDFCDEKRELSKVHAKRLSDERNLLQDPNPFKGINRQ 695
QY 676 MLQTFRASKLLDLRTRSRIFIPNGRTMMGCLDES-RLEYGVQV-----FVQFTGAGHG 727
Db 696 LDRGWGSTDTIQRGDDVFKENYVTLPGTFDECYPTLYQKIDESKKAFTRIQLRGYI 755
QY 728 EFSDDLHFNNSRSTNSNFIKLNWVAVKPNCLHPGDIRVLKAVNRALHHMVDVVFPQ 787
Db 756 EDSDDL-----EYILIRYN--AKH-----ETVNVPGTGLWPLSLAQSP 791
QY 788 KGKRPHPNECS-----GSDLD 803
Db 792 IGKCGEPNRCAPHLEWNPDL 812

RESULT 13

US-08-922-505A-10
; Sequence 10, Application US/08922505A
; Patent No. 6110464
; GENERAL INFORMATION:
; APPLICANT: Malvar, Thomas
; APPLICANT: Gilmer, Amy Jelen
; TITLE OF INVENTION: BROAD-SPECTRUM (-ENDOTOXINS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,505A
; FILING DATE: 03-SEP-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: MECO:163

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512)418-3000
TELEFAX: (512)474-7577
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-922-505A-10

Query Match 1.8%; Score 104; DB 3; Length 1177;
Best Local Similarity 18.2%; Pred. No. 1.2;
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FGCQISDDKFAVLGSTEVS-----IQFGIGLKKEFFFLSSGSADYKQLQSYENIMQVVLH 156
DB 13 YNC-LSNPEVVLGGERIETGYTIDISLTQFLSEFVPGAGFVLGL-VDIINGI--- 67
QY 157 RPYG--QNAQFL-----IQFGAPRIYKRL--NSCYSPFKETPDQWVRTTDFP 203
DB 68 --FGPSQWDAFLVQIQLNORIEEFARNOAISRLGLSNLIQIYAES-----PREWEAD 120
QY 204 PSWIGLSSLCLOPRGVRLENPEESFFHYAERENNITLOTGFTFFVSQKSALVPNQPP 263
DB 121 PTNPALREEMRIQF-----NDMNSALTTAIPLFA-----VQNTQVP 156
QY 264 -----EGISIPYKILFKISSLVQHGCIPGALN----- 291
DB 157 LLSVYVQAANLHLSVLRDVSFGQRMGFDAAATINSRYNDLTRLIGNYTDYAVRWYNTGLE 216
QY 292 -----VYFRLVDPRRNRVACIEHALEKLYIKKECCYD 324
DB 217 RVWGPDSDRWVRYNQFRRELTLVLDIVALFPNYDSRRYPITRVSQLTREIY-----TN 270
QY 325 PVRWLTEQDGYLKG-----RQPKSP-----SITL-----DDGLVY--VRRVLVTPCK 366
DB 271 PV---LENFDSGFRGSAQGIERSIRSHPLMDILNSIIYTDHARGYYIWSGHQIMASP-- 325
QY 367 VYFCGPEVNVSNRVLNRYSEDIDNPLRVSFVDEWEKLYST--DLLPK-----ASTGSGVR 420
DB 326 VFGSGPEFTF----- 468
QY 421 TNIVERILSTL-RKGFVIGDKKFEFLAFSSQLR--DNSVWMAFASRPLTA----- 468
DB 359 ---YRTLSTLYRRPNIG-----INNQLSVLDGTEFAYGTSNLPASVYRKSGTV 407
QY 469 -----NDIRAWMGDFSOIKNVAARLQSGFSSRETLVLR----- 506
DB 408 DSLDEIPQNNNVPYPRQGFSHRLSHVSMFRS-----GFSNSSVSIIRAPMFSWTHRSAT 461
QY 507 -----HEIEVIPDKVH-----GTSYVPSDGI--GKI-----SGDFAHRVASKCGLOYT 548
DB 462 PTNTIDPERITQIPLVKAHTLQSGTTVVRGPGFTGGDILRTSGGPFAYTIVNINGOLPQ 521
QY 549 PSAQIRYGG-----YKGVGV-----DPDSSMKSLRKMSKY----- 582
DB 522 YRARIRYASTNLRIYTVAGERIFAGQFNKMTDGLPTFFQSFYSATINTAFTEPMSQ 581
QY 583 ESDNKLVDLGWSKQPCYLN-----OLITLLSTLGVKDE 618
DB 582 SSFTVGADT--FSSGNEYVIDRFEIIPVTAITEAYDLERAQKAVNALFTSINGIGIKTD 639
QY 619 VLEQOKAEVQDLOAI-----LHDSLKAQBAELMSPGENTNLKAMLCNGYKPDAPPELSM 675
DB 640 VTDYH-----IDQVSNLVDCLSEDFCLDEKRELSEKVKHAKRLSDERNLLQDPNFKGINRQ 695
QY 676 MLQFTFRASKLLDTRSRIFTNRTMGCLDES-RILEYQV-----FVQFTGAGH 727
DB 696 LDRGWRGSTDITIQRGDDVFKENYVLTGTFDECPYLYOKIDESIKAFTRYQLRGYI 755
QY 728 EFSDDLHPFNNSRSTNSNFIKGNVNAKNPCLHPGDIRVLKAYNRALHHMVDVCFVFPQ 787

DB 756 EDSQDL-----EIVLIRYN---AKH-----ETVNVPGTGLWPLSAQSP 791

QY 788 KGRPHNECS-----GSDLD 803

DB 792 IGKCGEPNRCAPHLEWNPDL 812

RESULT 14

US-09-260-952A-10
Sequence 10, Application US/09260952A
Patent No. 6221649
GENERAL INFORMATION:
APPLICANT: Malvar, Thomas
APPLICANT: Gilmer, Amy Jelen
TITLE OF INVENTION: HYBRID BACILLUS THURINGIENSIS DELTA-ENDOTOXINS WITH
FILE OF INVENTION: NOVEL BROAD SPECTRUM INSECTICIDAL ACTIVITY
FILE REFERENCE: MECO:217
CURRENT APPLICATION NUMBER: US/09/260,952A
CURRENT FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 10
LENGTH: 1177
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: coding
US-09-260-952A-10

Query Match 1.8%; Score 104; DB 4; Length 1177;
Best Local Similarity 18.2%; Pred. No. 1.2;
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FGCQISDDKFAVLGSTEVS-----IQFGIGLKKEFFFLSSGSADYKQLQSYENIMQVVLH 156
DB 13 YNC-LSNPEVVLGGERIETGYTIDISLTQFLSEFVPGAGFVLGL-VDIINGI--- 67
QY 157 RPYG--QNAQFL-----IQFGAPRIYKRL--NSCYSPFKETPDQWVRTTDFP 203
DB 68 --FGPSQWDAFLVQIQLNORIEEFARNOAISRLGLSNLIQIYAES-----PREWEAD 120
QY 204 PSWIGLSSLCLOPRGVRLENPEESFFHYAERENNITLOTGFTFFVSQKSALVPNQPP 263
DB 121 PTNPALREEMRIQF-----NDMNSALTTAIPLFA-----VQNTQVP 156
QY 264 -----EGISIPYKILFKISSLVQHGCIPGALN----- 291
DB 157 LLSVYVQAANLHLSVLRDVSFGQRMGFDAAATINSRYNDLTRLIGNYTDYAVRWYNTGLE 216
QY 292 -----VYFRLVDPRRNRVACIEHALEKLYIKKECCYD 324
DB 217 RVWGPDSDRWVRYNQFRRELTLVLDIVALFPNYDSRRYPITRVSQLTREIY-----TN 270
QY 325 PVRWLTEQDGYLKG-----RQPKSP-----SITL-----DDGLVY--VRRVLVTPCK 366
DB 271 PV---LENFDSGFRGSAQGIERSIRSHPLMDILNSIIYTDHARGYYIWSGHQIMASP-- 325
QY 367 VYFCGPEVNVSNRVLNRYSEDIDNPLRVSFVDEWEKLYST--DLLPK-----ASTGSGVR 420
DB 326 VFGSGPEFTF----- 468
QY 421 TNIVERILSTL-RKGFVIGDKKFEFLAFSSQLR--DNSVWMAFASRPLTA----- 468
DB 359 ---YRTLSTLYRRPNIG-----INNQLSVLDGTEFAYGTSNLPASVYRKSGTV 407
QY 469 -----NDIRAWMGDFSOIKNVAARLQSGFSSRETLVLR----- 506
DB 408 DSLDEIPQNNNVPYPRQGFSHRLSHVSMFRS-----GFSNSSVSIIRAPMFSWTHRSAT 461
QY 507 -----HEIEVIPDKVH-----GTSYVPSDGI--GKI-----SGDFAHRVASKCGLOYT 548
DB 462 PTNTIDPERITQIPLVKAHTLQSGTTVVRGPGFTGGDILRTSGGPFAYTIVNINGOLPQ 521

QY 549 PSAQIRYGG-----YKGVVGV-----DPDSSMKLSLRKMSKY----- 582
DB 522 RYRARIYASTNRIYVTVAGERIFAGQFNKMTDGTPLTFQSFYATINTAFTFPMQ 581
QY 583 ESDNIKLDVLGWSKYQPCYLNR-----OLITLLSTLGVKDE 618
DB 582 SSFTVGADT--FSSGNEVYIDRFELIPVTATFEAYDLERAQKAVNALFTSINGIKTD 639
QY 619 VLEQKQKEAVDQDAI---LHDSLKAQEALELMSPGENTNLIKAMLCGKYPDAEPFLSM 675
DB 640 VTDYH---IDQVSNLVDCLDEFLDEKRELSKVHAKRLSDERNLLQDPNFKGINRQ 695
QY 676 MLOTFRASKLLDLRTRSRIFIPNGRTMMGCLDES--RTLEYGOV-----FVQFTGAGHG 727
DB 696 LDRGWRGSTDTIQRGDDVFKENYVTLPGTDECPYLYQKIDESKLKAFTRQYLRGI 755
QY 728 EFSDDLHPFNNSRSTNSNFIKNGVVAKNPCPLHPGDIRVLKAVNRALHHMDCVFPQ 787
DB 756 EDSQDL-----EYILIRYN---AKH-----ETVNVPGTGLWPLSAQSP 791
QY 788 KGKRPHPNECS-----GSDLD 803
DB 792 IGKCGEPNRCAPHLEWNPDL 812

RESULT 15

US-09-253-341-10
; Sequence 10, Application US/09253341
; Patent No. 6242241

GENERAL INFORMATION:

APPLICANT: Maivaer, Thomas
; Gilmer, Amy Jelen
; TITLE OF INVENTION: BROAD-SPECTRUM (-ENDOTOXINS)
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/253/341

FILING DATE: 10-Feb-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/922,505

FILING DATE: 03-SEP-1997

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: MECO:163

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512)418-3000

TELEFAX: (512)474-7577

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1177 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-09-253-341-10

Query Match

Best Local Similarity 1.8%; Score 104; DB 4; Length 1177;

Matches 168; Conservative 112; Mismatches 112; Indels 340; Gaps 45;

Search completed: November 6, 2002, 03:47:17

Job time : 28 secs

QY 102 FGQIISDDKFAVLGSTEVS-----IQGIGLKKFFFLSSGSADYKLLQLSYENIQWVLH 156
DB 13 YNC-LSNPEVEVLGGERIETGYTPIDISLTOFLSEFVPGAGFVLGL-VDIWGI--- 67
QY 157 RPYG--QNAQFL--TOLGAPRIYKRL--NSCYFFKRETPDDQVVRTDPP 203
DB 68 --FGPSQWDAFLVQIEQLINORIEEFARNQAIISRLGLESLNYQIYAES-----FREWEAD 120
QY 204 PSWIGLSSSLCLOFRGVRLPNPEESFFHYAERENITLTGTFTFFVSOKSALVPNVQPP 263
DB 121 PTNPALAEENRIQF-----NDMSALTATPLFA-----VQNVQVP 156
QY 264 -----EGISPIYKILFKISSLVQHGCIPGALN----- 291
DB 157 LLSVYVQAANLHSLVRDVSFGQRMGFDAAATINSRYNDLTRIGNYDYAVRWYNTGLE 216
QY 292 -----VYFFVLVDPRRRNVACIEHALEKLYIYKECCYD 324
DB 217 RVNGPDSRWVRYNQFRRELTPLVDIVALFPNYSRRYPIRVISQLTRIY-----TN 270
QY 325 PVRLWTEQDGYLKG-----RQPKSP-----SITL-----DDGLVY---VRRVLVTPCK 366
DB 271 PV---LENFDGSPRGSQAQIERSIRSPHLMILNSITITDAHRGYVYWSGHQIMASP-- 325
QY 367 VYFCGPEVNVSNRVLNRYSEDIDNFLRVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420
DB 326 VGFSGPPEFTF-----PLYGTMGNAAPQOQRIIVAQLOGCV- 358
QY 421 TNIYERILSTL-RKGFVIGDKKFEFLAFSSQLR--DNSVYMFASRPLFA----- 468
DB 359 --YRTLSSLTLYRFPNIG-----INNOQLSVLDGTEFAYGTSNLPSSAVYRKSGTV 407
QY 469 -----NDIRAWMGDFSQKNVAKYAAARLGQSSRETLISVLR----- 506
DB 408 DSLDEIPQNNVPPRQGFSHRLSHVSMFRS-----GFSNSVSVIIRAPMFWTHRSAT 461
QY 507 -----HELEVIPDVAVH---GTSYVFSGDI---GKI-----SGDFAHRVASCKGLQYT 548
DB 462 PTNIDPERITOILVKAHLOSCTTVVVGPGFTGGDILRRTSGGPFAYITVNINGQLPQ 521
QY 549 PSAQIRYGG-----YKGVVGV-----DPDSSMKLSLRKMSKY----- 582
DB 522 RYRARIYASTNRIYVTVAGERIFAGQFNKMTDGTPLTFQSFYATINTAFTFPMQ 581
QY 583 ESDNIKLDVLGWSKYQPCYLNR-----OLITLLSTLGVKDE 618
DB 582 SSFTVGADT--FSSGNEVYIDRFELIPVTATFEAYDLERAQKAVNALFTSINGIKTD 639
QY 619 VLEQKQKEAVDQDAI---LHDSLKAQEALELMSPGENTNLIKAMLCGKYPDAEPFLSM 675
DB 640 VTDYH---IDQVSNLVDCLDEFLDEKRELSKVHAKRLSDERNLLQDPNFKGINRQ 695
QY 676 MLOTFRASKLLDLRTRSRIFIPNGRTMMGCLDES--RTLEYGOV-----FVQFTGAGHG 727
DB 696 LDRGWRGSTDTIQRGDDVFKENYVTLPGTDECPYLYQKIDESKLKAFTRQYLRGI 755
QY 728 EFSDDLHPFNNSRSTNSNFIKNGVVAKNPCPLHPGDIRVLKAVNRALHHMDCVFPQ 787
DB 756 EDSQDL-----EYILIRYN---AKH-----ETVNVPGTGLWPLSAQSP 791
QY 788 KGKRPHPNECS-----GSDLD 803
DB 792 IGKCGEPNRCAPHLEWNPDL 812

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:45:23 ; Search time 13 Seconds

(without alignments)
1188.490 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVGPYLLSAEVK.....RPVLNLSRAQLSHRLVLK 1114

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 90412 seqs, 13869272 residues

Total number of hits satisfying chosen parameters: 90412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5864	100.0	1114	10	US-09-782-874-2
2	1182	20.2	218	10	US-09-782-874-3
3	112.5	1.9	1184	10	US-09-815-242-5229
4	112.5	1.9	1188	10	US-09-815-242-12125
5	109.5	1.9	1579	10	US-09-801-368-368
6	107	1.8	1357	10	US-09-815-242-11997
7	105.5	1.8	1080	10	US-09-781-558-4
8	105.5	1.8	1135	10	US-09-737-149-35
9	104.5	1.8	711	10	US-09-966-561-2
10	104	1.8	1177	10	US-09-873-873-10
11	104	1.8	1177	10	US-09-873-873-12
12	104	1.8	1177	10	US-09-873-873-14
13	103	1.8	1177	10	US-09-873-873-28
14	102	1.7	897	10	US-09-815-242-11609
15	102	1.7	1331	10	US-09-801-368-370
16	101.5	1.7	1115	10	US-09-781-558-2
17	100	1.7	1177	10	US-09-873-873-26
18	99.5	1.7	914	10	US-09-823-356-8
19	99.5	1.7	914	10	US-09-922-217-1066

20	99.5	1.7	914	10	US-09-833-263-1066
21	99.5	1.7	2813	10	US-09-381-261A-1
22	99	1.7	650	12	US-10-115-178-1
23	98.5	1.7	932	10	US-09-788-657-17
24	98.5	1.7	932	10	US-09-788-657-18
25	97	1.7	1019	10	US-09-434-066-23
26	96.5	1.6	624	10	US-09-815-242-5640
27	96.5	1.6	625	10	US-09-815-242-12704
28	96	1.6	1502	10	US-09-801-368-140
29	95.5	1.6	736	10	US-09-815-242-11293
30	95.5	1.6	892	10	US-09-203-658-42
31	95.5	1.6	892	10	US-09-844-353A-42
32	95.5	1.6	991	10	US-09-815-242-5803
33	95.5	1.6	5701	10	US-09-864-761-37319
34	95	1.6	17	10	US-09-782-874-12
35	95	1.6	595	10	US-09-925-297-604
36	95	1.6	869	10	US-09-815-242-10623
37	95	1.6	1216	10	US-09-815-242-13312
38	94.5	1.6	552	10	US-09-817-784-4
39	94.5	1.6	709	10	US-09-815-242-13349
40	94	1.6	935	10	US-09-757-781-1
41	94	1.6	1186	10	US-09-826-752-8
42	93.5	1.6	601	10	US-09-815-242-12358
43	93.5	1.6	1206	12	US-10-032-717-2
44	93.5	1.6	1216	10	US-09-815-242-13698
45	93.5	1.6	2813	10	US-09-886-900-2

ALIGNMENTS

RESULT 1

US-09-782-874-2

; Sequence 2, Application US/09782874

; Patent No. US20010023067A1

; GENERAL INFORMATION:

; APPLICANT: Wassenegeger, Michael

; Schiebel, Winfried

; Sanger, Heinz

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING

; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN

; RNA-DIRECTED RNA POLYMERASE (RRP)

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESS: RISH & NEAVE

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/782,874

; FILING DATE: 08-Feb-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/811,583

; FILING DATE: 05-MAR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Haley, James F.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: MPG-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-596-9000

; TELEFAX: 212-596-9090

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1114 amino acids

Query Match 100.0%; Score 5864; DB 10; Length 1114;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKIQVFPFYLISAEVWSFLEKTYGYGTVCALVCALEVKQSGSRFAKQVQFADNISADK 60
Db 1 MGKIQVFPFYLISAEVWSFLEKTYGYGTVCALVCALEVKQSGSRFAKQVQFADNISADK 60

QY 61 IITLANNRLYFGSSYLKAWENKTDIVQLRAYVDQMDGTTNFGCOISDDKFAVLGSTEVS 120
Db 61 IITLANNRLYFGSSYLKAWENKTDIVQLRAYVDQMDGTTNFGCOISDDKFAVLGSTEVS 120

QY 121 IOFGIGLKKFFFLSSGSADYKLOLSYENIWOVYLHRYPGONAQFLLIQLGAPRIYKRL 180
Db 121 IOFGIGLKKFFFLSSGSADYKLOLSYENIWOVYLHRYPGONAQFLLIQLGAPRIYKRL 180

QY 181 ENSCYSFKEPDDQWRTTDFPSSWIGLSSSLCQFRRGVRLNFEESEFFHYAERENNI 240
Db 181 ENSCYSFKEPDDQWRTTDFPSSWIGLSSSLCQFRRGVRLNFEESEFFHYAERENNI 240

QY 241 TLQGTFFVSKSALPNVOPPEGISIPYKILFKISLVOHGCIPOGALNVFFRLVDP 300
Db 241 TLQGTFFVSKSALPNVOPPEGISIPYKILFKISLVOHGCIPOGALNVFFRLVDP 300

QY 301 RRRNVACIEHALEKLYIKYKCCYDPVWLTEQYDYLKGRQPPKSPSITLDDGLVYVRRV 360
Db 301 RRRNVACIEHALEKLYIKYKCCYDPVWLTEQYDYLKGRQPPKSPSITLDDGLVYVRRV 360

QY 361 LVTCKVYFCGPEVNSNRVLRNYSIDNPLRVSFVDEWEKLYSTDLLPKASTGSGVR 420
Db 361 LVTCKVYFCGPEVNSNRVLRNYSIDNPLRVSFVDEWEKLYSTDLLPKASTGSGVR 420

QY 421 TNYIERILSTLRKGFVIGDKKFEFLAFSSQLRNSVWVFASRGLFANDIRAWMGDFSQ 480
Db 421 TNYIERILSTLRKGFVIGDKKFEFLAFSSQLRNSVWVFASRGLFANDIRAWMGDFSQ 480

QY 481 IKNVAKYAARLQSGFSGSRETLVLRHEIEVPVKVHGTSYVSDGIGKISGDFAHRA 540
Db 481 IKNVAKYAARLQSGFSGSRETLVLRHEIEVPVKVHGTSYVSDGIGKISGDFAHRA 540

QY 541 SKCGLQYTPSAFOIRYGYKGVGVDPSMKSLSRKSMSKYESDNKLDVLGWSKYQPC 600
Db 541 SKCGLQYTPSAFOIRYGYKGVGVDPSMKSLSRKSMSKYESDNKLDVLGWSKYQPC 600

QY 601 YLNROLITLLSTLGVKDEVLOKKEAVDQDLDAIHLDSLKAQEALELMSPGENTNILKAM 660
Db 601 YLNROLITLLSTLGVKDEVLOKKEAVDQDLDAIHLDSLKAQEALELMSPGENTNILKAM 660

QY 661 LNCGYKPAEPFLSMMLQTFRASKLLDTRSRIFIPNGRTMGMCLDESRTLEYGVFVQ 720
Db 661 LNCGYKPAEPFLSMMLQTFRASKLLDTRSRIFIPNGRTMGMCLDESRTLEYGVFVQ 720

QY 721 FTGAGHGFEFSDLLHFPNNSRSTNSFILKGNVYVAKNPLHGDIVRLKAVNVRALHHV 780
Db 721 FTGAGHGFEFSDLLHFPNNSRSTNSFILKGNVYVAKNPLHGDIVRLKAVNVRALHHV 780

QY 781 DCWVFPQKGRPHNPNCSGSLDGDYFVCWQDMIPPRQVQPMYPPAPSIQDHDVTI 840
Db 781 DCWVFPQKGRPHNPNCSGSLDGDYFVCWQDMIPPRQVQPMYPPAPSIQDHDVTI 840

QY 841 EEVEEFTNYIVNDSIGLIIANAHVVADREPDMASDPCKKLAELFSTAVDFPKTGPAP 900
Db 841 EEVEEFTNYIVNDSIGLIIANAHVVADREPDMASDPCKKLAELFSTAVDFPKTGPAP 900

QY 901 IPSQLRPKEYPDMKPKDTSYISERVIGKLFVRKVKAPQASSIATFTTRDVARRSYDAD 960
Db 901 IPSQLRPKEYPDMKPKDTSYISERVIGKLFVRKVKAPQASSIATFTTRDVARRSYDAD 960

QY 961 MEVDGFEDYIDAFDYKTEYDNKLGNDYGYGKTEAEILSGGIMKASKTFDRRKDAEAI 1020
Db 961 MEVDGFEDYIDAFDYKTEYDNKLGNDYGYGKTEAEILSGGIMKASKTFDRRKDAEAI 1020

QY 1021 SVAVRALRKEARAWFKRRNDIDDMPLKASAWHYVTHPTYWGYNOLKRAHFISFPWCV 1080
Db 1021 SVAVRALRKEARAWFKRRNDIDDMPLKASAWHYVTHPTYWGYNOLKRAHFISFPWCV 1080

QY 1081 YDOLIOIKDKKARNRPVNLSSLRALSHRLVLK 1114
Db 1081 YDOLIOIKDKKARNRPVNLSSLRALSHRLVLK 1114

RESULT 2
US-09-782-874-3
: Sequence 3, Application US/09782874
: Patent No. US20010023067A1
: GENERAL INFORMATION:
: APPLICANT: Wassenegger, Michael
: Riedel, Leonhard
: Schiebel, Winfried
: Sanger, Heinz
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
: POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN
: RNA-DIRECTED RNA POLYMERASE (RdRP)
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FISH & NEAVE
: STREET: 1251 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10020
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA: US/09/782,874
: APPLICATION NUMBER: US/09/782,874
: FILING DATE: 08-Feb-2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/811,583
: FILING DATE: 05-MAR-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Haley, James F.
: REGISTRATION NUMBER: 27,794
: REFERENCE/DOCKET NUMBER: MPG-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-596-9000
: TELEFAX: 212-596-9090
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 218 amino acids
: TYPE: amino acid
: STRANDEDNESS: <Unknown>
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-782-874-3

Query Match 20.2%; Score 1182; DB 10; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.1e-100;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 RTMGMCLDESRTLEYGVQVFTGAGHGFEFSDLLHFPNNSRSTNSFILKGNVYVAKNPC 759
Db 1 RTMGMCLDESRTLEYGVQVFTGAGHGFEFSDLLHFPNNSRSTNSFILKGNVYVAKNPC 60

QY 760 LHPGDIRVLKAVNVRALHHVDCVFPQKGRPHNPNCSGSLDGDYFVCWQDMIPPR 819
Db 760 LHPGDIRVLKAVNVRALHHVDCVFPQKGRPHNPNCSGSLDGDYFVCWQDMIPPR 819

```
Db 61 LHPGDIRVLKAVNRALHMHVDCVVPQKGRKPHNECSGSDLDGDIYFVCWDQDMPPR 120
QY 820 QVQMEYPPASIQLDHVDVTEEVETNYIVNDSLGIITANAHVVFADREPDNMSDPC 879
Db 121 QVQMEYPPASIQLDHVDVTEEVETNYIVNDSLGIITANAHVVFADREPDNMSDPC 180
QY 880 KKLAEFLSIADVFRTGVPABIPSQLRPKEYPDMFMDKP 917
Db 181 KKLAEFLSIADVFRTGVPABIPSQLRPKEYPDMFMDKP 218

RESULT 3
US-09-815-242-5229
; Sequence 5229, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5229
; TYPE: PR
; ORGANISM: Staphylococcus aureus
US-09-815-242-5229

Query Match 1.9%; Score 112.5; DB 10; Length 1184;
Best Local Similarity 18.2%; Pred No. 0.28;
Matches 162; Conservative 124; Mismatches 299; Indels 305; Gaps 40;

QY 324 DPVRM-LTEQYDGYLKGQPPKSPSITLDGLVYVRRVLVTPCKYFCGPEVNSNRLR 382
Db 42 DAIKWVLGQSASLRSK------MED-----IIFSGAE-----HRKAQ 75
QY 383 NYSE---DIDNPLRVSVFDEWEKLYSDLLPKASTGSGVRTNIVERILSLRKGFLVD 439
Db 76 NYAEVQLRDNHNSKLSVDE-----NEVIVT-RDLIRSGE 109
QY 440 KKEFLAFSSQLRNSNVMFASRPGLTANDIRAMWGFDSQTKN-----V 484
Db 110 SEY-VINNDRAKLDIADFLDSGLGKEAYSIIIS-QGRVDEILNAPKIDRRQIIESAGV 167
QY 485 AKYARLGQSGSSRETLSVLRIHEVTPDVKVHGTSTVFSDGIGKISGDFAHRVASKG 544
Db 168 LKYYKRAESLNKLQDTEDLNLRVEDILYDLE-----GRVEPLKEEAIAIAK-- 213
QY 545 LQYTSFAOIRYGGYGVGVVDPDSSMKLSLRKSMKSYESDNKLDVLGWSKYQPCYLN 604
```

```
Db 214 -EYKTLISHQMKHSDI--VVTV-----HDIDQYTNDRQLD----- 245
QY 605 QLITLLSTLGVKDEVLEQKQKEAVDO-----LDAILHDSLKAQEALELMSPGEN 653
Db 246 QRLNDLQGOQANKEADKQRLSQIQYKGRKHOLDNDVSNLYQVKAFAFEKYI--GQ 303
QY 654 TNIL-----KAMLNGYKPDAPFELSM-----LQTRASKLLDLTRSRIFIP 697
Db 304 LNVLEERKKNQSETNARYEEQENLIELLENISNEASEADTYKSLKSKQKELNAVI-- 360
QY 698 NGRMTMGCLDESRTLEGVGVQFTGAGHGEFSDDLHPFNNSRSTNSNFTLKGNVVAKN 757
Db 361 -----RELE-EQDIY-----SDEAHDEKLEIEIKNEYITLMEQSDVNN 397
QY 758 POLHPGDIRVLK-----AVNRALHMHVDCVVPQ-----KGRKPHNECS 798
Db 398 -----DIRFLKHTIENEAKSKRLDSRLVE--VFQQLKDIOQGIKTKKEYQOTNKELS 449
QY 799 GSD-----LDGDIYFVCWDQDMPPRQVQPMYPPAPSIQLDHVDVTEEVETNYI-- 851
Db 450 AVDKREIKNIEKDLTDTKKAQNEYEEKLYQAYRTEKMKTRIDSLATQEEETVFFNGVKH 509
QY 852 ---VNDSLGIIANAHVVFADREPDNMSDPCPKKLAEFLSIADVFRTGVPABIPSQLRP 907
Db 510 ILKAKNKELKGIHGAVAEIIDVPSKLTQATETALGASLQHVIVDSEKDGQA----- 561
QY 908 KEYPDFMDKPKDTSYISERVIGK-LPRKYKDKAPQASSINATFTEDVARRSYDADMEVDGF 966
Db 562 -----IQFLKERNLGRATEFLPL--NVIOQRVAVDIDIKSIK-----EANGF 600
QY 967 EDYIDEAFDYKTEVDNKLGNLM-----DYY-----GIKTEAEI----- 999
Db 601 ISASEAVKVAPEYQNIIGNLIGNTIIVDLKHANELARAIVKTRIVTLEGDIVNPGGS 660
QY 1000 LSGIMKASTFDRRKDAEAIISAVRALRKEARAKWRKRNDDIDMLPKASAWHVYHPT 1059
Db 661 MTGGARKSKSILSKD-ELTTM-----RHQLEDYLQRTES----- 695
QY 1060 YMGYNQGLKRAHFISPPWCVIDLIQIKKDKARNRPVNLNLSLRAQLSH 1109
Db 696 ----FEQPFKELKIKS-----DQLSELYFEKSKQH-----NLTKEQVHH 730

RESULT 4
US-09-815-242-12125
; Sequence 12125, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
```

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12125
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12125

Query Match 1.9%; Score 112.5; DB 10; Length 1188;
Best Local Similarity 19.2%; Pred. No. 0.84;
Matches 162; Conservative 124; Mismatches 299; Indels 305; Gaps 40;
QY 324 DPVRM-LTEQDGVKGRQPKSPSITLDDGLVTVRRVLVTPCKVYFCGPVNVNRVL 382
DB 43 DAIKWVLGEQSAKSLRGSK-----MED-----IIFSGAE---HRKAQ 76
QY 383 NYSE---DIDNPLRVSVFDEWEKLYSTDLPLPKASTGSGVTRNIYERILSTLRKGFVIGD 439
DB 77 NYAEVQLRLDNRHKKLSYDE-----NEVIVT-RLYRSGE 110
QY 440 KKFEPLAFSSQOLRNSVWVFASRPLGTANDIRAWMGDFSOIKN-----V 484
DB 111 SEY-YINDRARLKDIALDLDSGLGRKAYSIIIS-QGRVDEILNAKPIDRQIIIESAGV 168
QY 485 AKYAARLQCSGSRBETLSVLRHIEVIPDVKVGTSYVSDGKIGSGDFAHRVASKCG 544
DB 169 LKYKRKAESLKNLQDTECNTRVEDILYDLE-----GRVEPLKEEAIAK-- 214
QY 545 LQYTPSAFQIRYGGYGVGVDPSSMKLSLRKSMKSYESDNIKDLVLGWSKYQPCYLNR 604
DB 215 -EYKTLHQMKHSDI--VVTV-----HDIQYTNDRQLD----- 245
QY 605 QLITLLSLTGLVKDEVLQKQKAVDQ-----LDAILHSLKAQEAELMSFGEN 653
DB 247 QRLNDLQOQAKKEADKORLSQIQYKGRHQLDNVSNLYQVLKATEAFERYT--GQ 304
QY 654 TNIL-----KAMLNCGYKPAEPFLSNM-----LQTFRASKLLDLRTSRIFIP 697
DB 305 LNVLEERKNGSETNARVEEQENLIELLENISNEISQAQTYSLKSKQKELNARI-- 361
QY 698 NGRTWMCGLDSRTLEYQGVFVQTAGHGFSDDLHPFNNSRTSNFILKGNVYVAKN 757
DB 362 -----RELE-EQLYV-----SDEAHKELEIKNEYITLMSQSDVNN 398
QY 758 PCLHPGDIIRVLK-----AVNRALHHWDCVVFPPQ-----CKRHPNECS 798
DB 399 -----DIRLUKHTTEENAKKSLDSRLVE--VFQKLDIQOIKTKKEYQOTNKELS 450
QY 799 GSD-----LDGDIYFVCWDQMIPRQVQPMYPPAPSIQLDHDVTIEVEVEYFTNYI-- 851
DB 451 AVDKKIKNIENLDLTKKAQNEYEKLYQAYRYTEKMKTRIDSLATQEEYTYFFNGVKH 510
QY 852 ----VNDLSGLIIRAHVFADEPDMSDCKKLAEFLSFSTAVDFPKTGVPAEIPSLRP 907
DB 511 ILKANKKELKGHGAVAEIIIDVPSKLTQAIETALGASIQHVIVDSEKDGQA----- 562
QY 908 KEYPDFMDPKPQTSYISERVIK-LFRKVKDKAPQASSIATFTDVARRSYDADMEVDGF 966
DB 563 -----IQPLKERNLGRATFLPL--NVIQSRVAVTDIKSIK-----EANGF 601
QY 967 EYIDAEADYKTEYDNKLGNUM-----DYY-----GKTEAEI----- 999
DB 602 ISIASEAVKVAPEYQNIIGNLGNITIIYDHLKHANELARAIRYTRIVTLEGDIVNPGGS 661
QY 1000 LSGGTMKASKYFDRDKDAEATISVAVRALRKEARWFKRRNDIDMDLPKASAWYHVYHPT 1059
DB 662 MFGCGARKSKSLSKD-ELTVM-----RHOLEDYLRGTES----- 696
QY 1060 YWCGYNOGLKRAHFTSFPCWYVDQIIQIKDKARNRPVLNLSRLAQOLSH 1109
DB 697 ----FEQGFKEKLIKIS-----DQLSLEYFEKSKQH-----NLTKEQVHH 731

RESULT 5

US-09-801-368-368
; Sequence 368, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; FILE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 368
; LENGTH: 1579
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-368

Query Match 1.9%; Score 109.5; DB 10; Length 1579;
Best Local Similarity 19.2%; Pred. No. 0.84;
Matches 247; Conservative 184; Mismatches 447; Indels 409; Gaps 73;
QY 15 SAEVKSFLKYGTYGTVCALEVYKQSGSRAPAKVQFADNIS---ADKIIITLA---NNR 68
DB 474 AAEIKSLVDKY-----YRVNLNMPNKRMAERKPIITTEAFRRNR 513
QY 69 LYFGSKYLYK-AWEMKTDIVOLRAVY-----DQMDGITLN-----FGCQ 105
DB 514 IDTLNSKWLNFKNFEDINIAVLKKWIVGNKELESTTEVDNTVNLDDPAVFATNCKRFAEQ 573
QY 106 ISDDKFAVLGSTEVSQFGIG---LKKFFFLSSGSADYKLQLSY-ENWQVVLHRPYGQ 161
DB 574 IMKEKDIEL-IFOKKIFFPLAPWILKAKFFFLKYQKTNELNLSYLDQDLEFLLMFPMRL 632
QY 162 NAQFLIQLFGAPRIYKRLNESCYSFFKETPPDQWVRTDPPSPWIGLS-----SSLC 214
DB 633 VKDIIILRLSYA---KKIQNPILMIDQMMDDF-----STYIKLAVQMKFTVASYC 680
QY 215 LQFRGVRL-PNEE-----ESFFHYAERENNITLOTGTFVFSOKSALVPWQPEGI 266
DB 681 NDWFFKVIDPEDHTVVEGLEYYFSILE-----LRILYSKNSFKTSKEP--DL 728
QY 267 SIYPKILF-----KISSLYOHGCIPOGALNVYFFRL--VDPRRRV 305
DB 729 LLYKWFNRVNGYIIDDAGELIAAEFTKLTLRVH-----RLHAYLLRQONTPKLENE 782
QY 306 ACIEHALEKUYI-----KECCYDPVRNLTEQYD----- 334
DB 783 AAAAEKVLQVFEILGSMKRLNFTNLTAKFQNFVRYKIEDHNYLLKQKETHFLIYT 842
QY 335 -GYLK--GRQPPKSPSI--TLDDGLVTVRRVLVTPCKVYFCGPVNVNRVLRNYSIEDID 389
DB 843 GGLEQNGTYLIGSPPELLGCKDDDIIRIKNSDIGDLV---PKLEINN-----SLTIY 893
QY 390 NFLRVSVDEWEKLYSTDLPLPKASTGSGYR---TNIY---ERILSTLRKGFVIGDKKFEF 444

```
Db 894 NAL-----DDWNNS-----NSLSDISNDGTPPYIKNDITTOPRSY-NGNRVNR 939
Qy 445 LAFSSQRLDNRVWFMASRPG-----LTANDIRAWMGDFSQIKNVAKYAARLGQSGFS 497
Db 940 PDFENSRSTEEFELEFELNSGLVVLITQEPILLWEGEMYNL-----S 984
Qy 498 SREILSVLRHIEVIP-----DVKVIGTSVFS-----DGKIGKISG-----DFAHRVAS-----KCG 544
Db 985 DNKTIKPEGLNLKIPNSIDLKCGSSYALEYQCDRFQOISGSSVSFELEKSSSETYKNN 1044
Qy 545 LOYTPSA-FOIRYG-----CYKGWV-----GVDRDSSMKLS-----LRKSMKSYES 584
Db 1045 LORINKAYFCTYSVLKNYTKIVTFKYSVNDLLNIFLFGDRDGLNFLRINVANNEK 1104
Qy 585 DNKLDV-----LWGSKY-OPCYLNRLI-----TLSTGLGVKDEV 619
Db 1105 RSIILLMRLSIGWLKPLAECDCPTDQRFVRCWCVTSMFAHMHVSGWNILALDECQFSS 1164
Qy 620 LEOKOKEAVDOLDALDHLDSLRQAEALMSGENTNILKMLNCGYKPDAPFL--SMML 677
Db 1165 LKQISECMSSL--ISHFDIIGARSIEV-----EKIN-QQARSNL-----DLEDVDDMMML 1213
Qy 678 QT---FRASKLIDLRTSRIFPNRGTMMGCLDESRTLEYQGVFQVQFTGAGHGFSDDLH 734
Db 1214 QVNSEFRVQSIWELEERIK---RNPHTQGVKVIDDS---DKGNKYLVSLASSISNVVS--- 1263
Qy 735 PFNSRSTNSNFI---LGNVWVAKNCPCHPGDIRVLKAVNVRALHHMVDVFPQKGR 791
Db 1264 ---NRWOKRPFPGGTFGRVYSAVD---LDNGEILAVKEINIQDSKSMQK---IFPLIKEE 1315
Qy 792 ---PHPNecs---GSDLDGDYFVGCWDDMIPPROVQPMYPPAPPSIO---LDHD-VT 839
Db 1316 MSVLILNHPNIVSYGVEVHRDKVNF-----MEYCEGSLAALLEHGRIE 1362
Qy 840 IEVEEYFTNTVNDLSGIIANAHVVPADREPDMAMSDPCCKLAELFSAIVDFPKTGUPA 899
Db 1363 DEMVQVYTLQLL-EGLAYLHESGIVHRDKPE-----NILDF--NGV-- 1403
Qy 900 EIPSQLRPEKYPDFMDKPKDTSYISERVIGLFRKVKDKAPQASSIATFTTRDVARRSYDA 959
Db 1404 ---IKYVDF-----GAKKIANNGTFLASMKIE----- 1429
Qy 960 DMEVDGFEDYIDEAPDYKTEYDNKLNLMYYG---IKTEAILSGIMKASKTFDRKDA 1017
Db 1430 ---NADGEHEDVTHVSDSRKAVXNNE-NALLDMGTPMYMAPESITGSTTKGLGAD---DV 1483
Qy 1018 FAISVAVRALRKEARAFKRRNDIDDLPLKASANVHTVHYTHYHPCYNGQL----- 1068
Db 1484 WSLGCVLEMITGRPWANLONENALMYHVAAG--HTPQFPTKDEVSSAGMKFLERCLIQ 1541
Qy 1069 ---KRAHFISF---PWCYVDOLIQTK 1089
Db 1542 NPSKRASAVELLMDPW-----IVQIRE 1563
```

RESULT 6

```
US-09-815-242-11997
; Sequence 11997, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
```

```
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11997
; LENGTH: 1357
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-11997
```

```
Query Match 1.8%; Score 107; DB 10; Length 1357;
Best Local Similarity 18.9%; Pred. No. 1.1;
Matches 156; Conservative 107; Mismatches 268; Indels 296; Gaps 40;
```

```
Qy 331 EYD-----GYLK-----GROPKSPSTLDDGLVYVRRVLVTPCKYFCGPEVNVSNRVL 381
Db 398 ERYDLSAVGKMKNNRRIRGTETEGVGLSKEDIIDVLKTL-----VDIRNG-- 443
Qy 382 RNYSESIDNPF--LRVSFVDEWEKLYSTDLKPKASTGSGVYRTNIYBRILSTLRKGFVIGD 439
Db 444 KGIVDDIDHLGNRRVRCVGEMAEQFRVGLV-----RVERAVKERL----- 484
Qy 440 KKEFFAFSSQLRDNSSVWFASFAPGLTANDIRAWMGDFSQIKNVAKYAARLGQSGSSR 499
Db 485 -----SMAESEGLMPQD-----LINAKPVAIAKEFFGSSQ 515
Qy 500 -----ETLSVLRIEIVIPDKVHGTSYFSDGIGKISGDFAHRVASKCGIQVTPSA 551
Db 516 LSQFMDQNNFLSEITHKRRV-----SALGPGGLTERRAG 549
Qy 552 FQIR-----YGGYKGVGVVDPDSSMKLSRKMSKYESNKLKLDVLGWSKYQPCYLNQ 605
Db 550 FEVRDVHPTHYG--RVCPIETEPGNIGLINSIATYARTN-KYGFLE-SPYR--VVKDS 602
Qy 606 LIT-----LLSTLGVKDEVLEQ-----KQKEAVDOLDAILH-----DSLKAQEALELM--SP 650
Db 603 LVTDIEVFLSAIEADHIVIAQASATLNEKQLVDELVAVRHLNEFVYKAPEDVTLMDVSP 662
Qy 651 GENTNILKMLNCGYKPDAPFLS-----MMLOTFRASKLL-----DL 688
Db 663 QKVYVVAASLI-----PFLEHDDANRALMGSNMQROAVPTLRADKPLVGTGMERNV 713
Qy 589 RTRSRIFPNGR-TMCCLDSESRLEYGVFQVQFTGAGHGFSDDLHPFNNSRSTNSN-- 745
Db 714 ARDSGVCVWARRGVIDSDVASRVV-----VRVADDEVETGEAGVDI--YNLTKYTSRQN 767
Qy 746 -----FILKGNVW-----VAKNCPCLHPGDIRVLKAVNVRALHHMVDVYVFFQKGRPH 793
Db 768 TCINQRLVSKGDVARGDILADGSPSTMDEL-----ALGQNMVRVAFMPWNGFNFE 818
Qy 794 PNECSGDLGDYFVGCWDDMIPPROVQPMYPPAPPSIQLDHVDVTEEEVEEYFTNVI 853
Db 819 DS-----ICLSERVVOEDRFTTHIQELTVCVARDTKLGEETIADIPN-VGE 864
Qy 854 DSLGIIANAHVVPADREPDMAMSDPCCKLAELFSAIVDFPKTGVPAPRIPQLRPFKEPDF 913
Db 865 AALNKLEAGIYVGAEVQ-----AGDILVGKVTPKGEQLTPEE----- 904
Qy 914 MDPDKTYSYISERVIGLFRKVKD---KAPQAS-----STATTRDVARRSYDA---DM 961
```

```
Db 905 -----KLLRAIFGEKASDVKDTLSRVPTGKGVIDVQVETRGVERDSRALSIEKM 956
QY 962 EVDGFEDYIDEADYKTEYDNKGLNLMYYGKTEAEILSGIMKASKYDFDRKDA--EA 1019
Db 957 QLDQIRKDLNEEF-----RIVEGA-----IFERLRAALVGA 987
QY 1020 ISVAVRALRK-----EAAWFKRR---NDIDDMPLKASAW 1051
Db 988 KAEGGPALKRGKTEITDDYDLGLERGQWFKLRMADDAALNEOLEKAQAY 1034

RESULT 7
US-09-781-558-4
; Sequence 4, Application US/09781558
; Patent No. US20020034778A1
; GENERAL INFORMATION:
; APPLICANT: Beasley et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000756
; CURRENT APPLICATION NUMBER: US/09/781,558
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/199,811
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/641,426
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1080
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-09-781-558-4

Query Match 1.8%; Score 105.5; DB 10; Length 1080;
Best Local Similarity 18.5%; Pred. No. 1;
Matches 135; Conservative 102; Mismatches 249; Indels 245; Gaps 37;

QY 55 NISADKIITLANNRLYFGSSYLKAWEMKTDIVOLRAYVD-QMDGI-----TLNFGCOIS 107
Db 249 NIT-DFLLLTENNKFHLESVINITANLSSTKDLLSFLOQVMDNIRNSTTVMWFGC--- 304
QY 108 DDKFAVLGSTEVSIOFGIGLKKFFFTSSGSADYKLO-----LSYENIWQVVLHRPY 159
Db 305 -DMSIRQIFEMSTQGLSPPELHWLGDSQVBEELRTEGLPLGLIAHGKTTQSVFEY-Y 362
QY 160 GQNAQFLLIQLFG-APRIYKRLNSCYSPFKETPDQWVTTDFP-----PSWI 207
Db 363 VQDAMELVARAVATATMIQPEL-----ALLPSTMNCMDVKITNTLSQYLSRFLANTTFR 417
QY 208 GLSSSLCLOFRGRVRLPNFEESFFHYAERENNI--TLQGTFTFFVSQKSALVPNVQPEG 265
Db 418 GLSGSIKV---KGSTIISSENNFTWNLQHPMGKPMWTRLGSKQGGRIVMDSGINWEQA 474
QY 266 -----ISIPKILFKISSLVQH-----GCIPGPAALVYFFELVDPRRNVACI 308
Db 475 QRKHTFHQPNKHLRVLTLEHPFVFTREVDDGLCPAGOL-----CLDPMWNTDSSML 528
QY 309 EHALEKLY-----YIKECY-----QYDGYLKGKGRQPPKSPSITLDD 352
Db 529 DELFSSLSSNDYPIPKKCCYCYCIDLLEQLAEDMNFDFDLIVG-----D 576
QY 353 G-----LVVVRVRLVTPCKVY-----FCGPEVNVSRVL---RNY 384
Db 577 GKYGAWKNHWTGLVGDLLSGTANMAVTSFSINTARSOVIDFTSPFFSTSLGILVRATDT 636
QY 385 SEDIDNEL-----RVSFVDEWEKLYSTDLLPKASTGSGVATNIYER 426
Db 637 RAPICGAWPLHWTMWLQIFVALHITAIPLFLYEKWS-----PFGMTPKGRNRN--- 685
QY 427 ILSTLRKGFVIGDKKFEFLAFSSQLRDNYSYMFASRPLTA--NDIRAMWDFDSQTKN- 483
```

```
Db 686 -----KVFSF-----SSALNVCYALLF-----GRTAAIKPKKCTWGRF--LNNL 722
QY 484 -----VAKYAARL-----GOSFGSSRRTLS----- 503
Db 723 WAIFCMFCLSTYTANLAAMVGEKIYEELSGIHDPKLLHHPSQGFRFGTVRESSAEDYVRQ 782
QY 504 -----VLRHEIEVIPDKVHGTSYVFSOG-----IGKISGDFAHRYASKGLQYT 548
Db 783 SPENMEHYNRRYNVATPD-----GVQYLNKDEPKLDAIMDKALLDYEVSIDADCKLITV 838
QY 549 PSAFOIR-YGGYKGVGVDPDSSMKLSLRKSKSYESNIKLDYL--GWSKYQPCYLNQ 605
Db 839 GKPFALIEGYG-----IGLPPNSPLTSNISLISQYKSHGF-MDVLHDKWKYVVPCKGRSF 892
QY 606 LITLLSTLGVK 616
Db 893 AVTETLQMGIK 903

RESULT 8
US-09-737-149-35
; Sequence 35, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1135
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-09-737-149-35

Query Match 1.8%; Score 105.5; DB 10; Length 1135;
Best Local Similarity 18.5%; Pred. No. 1;
Matches 135; Conservative 102; Mismatches 249; Indels 245; Gaps 37;

QY 55 NISADKIITLANNRLYFGSSYLKAWEMKTDIVOLRAYVD-QMDGI-----TLNFGCOIS 107
Db 275 NIT-DFLLLTENNKFHLESVINITANLSSTKDLLSFLOQVMDNIRNSTTVMWFGC--- 330
QY 108 DDKFAVLGSTEVSIOFGIGLKKFFFTSSGSADYKLO-----LSYENIWQVVLHRPY 159
Db 331 -DMSIRQIFEMSTQGLSPPELHWLGDSQVBEELRTEGLPLGLIAHGKTTQSVFEY-Y 388
QY 160 GQNAQFLLIQLFG-APRIYKRLNSCYSPFKETPDQWVTTDFP-----PSWI 207
Db 389 VQDAMELVARAVATATMIQPEL-----ALLPSTMNCMDVKITNTLSQYLSRFLANTTFR 443
QY 208 GLSSSLCLOFRGRVRLPNFEESFFHYAERENNI--TLQGTFTFFVSQKSALVPNVQPEG 265
```



```
Db 444 GLSGSIK---KGSTIISSNNFFIWNLQHPMGKPMWIRLGSWQGRIVMDSGIWPQEA 500
QY 266 -----ISIPKILFKISSLVQH-----GCIPGALNVYFRLVDPRRNVACI 308
Db 501 QRKTHQHPNKLHLRVTLIEHFVTRVDDGLCPAGOL-----CLDPMNDSSML 554
QY 309 EHAEKLY-----YKECCY-----DPVRWLTE-----QYDGLKGRQPPKSPSITLDD 352
Db 555 DRFLSSLHSSNDTPVPIFKKCCYCYCIDLLEQLAEDNFDLYIVG-----D 602
QY 353 G-----LVYVRVLVTPCKY-----FCGPEVNVNRLV-----RNY 384
Db 603 GYGAWXNGHWGLVGLLSTANNAVTSINTARSQVIDFSPFFSTSLGILVTRDT 662
QY 385 SEDIDNEL-----RVSFVDEWEKLYSTDLLPKASTGSGVRTNIYER 426
Db 663 AAPIGAFMWPLHWMLGIFVALHITAFILTYEKS-----PFGMTFKGRNR-----711
QY 427 ILSTLRKGFVIGDKKFEFLAFSSQLRDNVWMPASRGLTA--NDRAMWGDFPSQIKN- 483
Db 712 -----KVFSF-----SSALNVCYALLF-----GRTAAKPKPCWTGRF--LMNL 748
QY 484 -----VAKYAARL-----GOSFGSSRETLs-----503
Db 749 WAIFCMCLSTYPTANLAAMVGEKIYELSGIHDPKLLHPSQGRFGTVRESSAEDYVRQ 808
QY 504 -----VLRHEVTEIPDVKHGTSYVFSDG-----IGKISGDPFAHRVASKCGLQYT 548
Db 809 SFPMEHYMYRRYVNPATPD-----GVQYKNDPKLDAFIMDKALLDYEVSIDACKLLTV 864
QY 549 PSAFOIR-YGYKGVGVGDPSMKLSLRKSMKSYESDNIKLDVL--GWSKYOPCYLNRQ 605
Db 865 GKPAIEGYC-----IGLPPNSPLTSMISELSQYKSHGP-MDLVLDKDKYKVPVCKRRSF 918
QY 606 LITLLTLGVK 616
Db 919 AVTETLQMGK 929

RESULT 9
US-09-966-561-2
; Sequence 2, Application US/09966561
; Patent No. US2002090696A1
; GENERAL INFORMATION:
; APPLICANT: University of Southern California
; APPLICANT: Miller, Carol A.
; APPLICANT: Dong, Zhao Hui
; APPLICANT: Zhang, Yan
; FILE OF INVENTION: APOPTOSIS INHIBITION
; TITLE REFERENCE: 13761-724
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/09/966, 561
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/419, 694
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-561-2

Query Match 1.8%; Score 104.5; DB 10; Length 711;
Best Local Similarity 20.2%; Pred. No. 0.62;
Matches 53; Conservative 40; Mismatches 99; Indels 71; Gaps 13;

QY 781 DCVVPKQKRPHPNECSGDLGGDIFVVCWDQDMTPPRVQPM-----EYPPAPS 831
Db 314 DPAAIPSTAGRPHPS---ISEEEG---FDC-----LSPERAEPGGGWRGSLGEPPPP 364
QY 832 IQLDHDTIEBEVEYFTNYIVNDSLGIIANAHV-----FADREPDMAMSDPCKKLA 883
Db 365 ASLSSDTALSVDVKYTLVVDH-----AQLELVSLRCPGSDYSDESDSATVYDNCASVS 420
```

```
QY 884 ELFSIAV-----DFPKTGYPAEIPSLRKEYPDFDKDKPKTSYISERVIGLFRKVKDK 938
Db 421 SPYESAIGEYEAAPRPQPPACLSEDSTP-----DEPD--VHFSKKFL-NVFMGRGR 470
QY 939 APOASSIAFT-----RDVARRSYDADMEVDG-----FEDYIDEAFD 975
Db 471 SSSAESFGLFSCIINGEEOQTHRAIFRVPRHDEDELEVEDDPLIVE-CAEDYWEAYN 530
QY 976 YKTEYDNKLGMLMDYIGIKTEAE 998
Db 531 MRT---GARGVFPAYIAIEVTKE 550

RESULT 10
US-09-873-873-10
; Sequence 10, Application US/09873873
; Patent No. US20020064865A1
; GENERAL INFORMATION:
; APPLICANT: Malvar, Thomas
; APPLICANT: Gilmer, Amy Jelen
; FILE OF INVENTION: Polynucleotide Compositions Encoding Broad-Spectrum S-Endotoxin
; TITLE REFERENCE: MECO:210--2
; CURRENT APPLICATION NUMBER: US/09/873, 873
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: US 09/253,341
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: US 08/922,505
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: US 08/754,490
; PRIOR FILING DATE: 1996-11-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hybrid Delta-Endotoxin
US-09-873-873-10

Query Match 1.8%; Score 104; DB 10; Length 1177;
Best Local Similarity 18.2%; Pred. No. 1.6;
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FCCQISDDRFVAVLGSTEVs-----IQFGIGLKKFFFFSSGADYKQLQLSYENIMQVILH 156
Db 13 YNC-LSNPEVEVLGGRIETGTPIDISLQFLISEFVPGAGFVLGL-VDIIMG- 67
QY 157 RPYG--QNAQFL-----IQLFGAPRIYKRLE--NSCYSEFFKETPDQDQVWRTDPP 203
Db 68 --FGPSQWDAFLVQIEQLINQRIEFARNOAISRLGSLNLYQIYAES-----FREWEAD 120
QY 204 PSWIGLSSSLCQFRGRVPLNPFESFFHYABRENNITLQGTFTFFVSQKSALVPNVQPP 263
Db 121 PTNPALREEMRIQF-----NDMNSALTATPLFA-----VQNYQVP 156
QY 264 -----EGISIPYKILFKISSLVQHCCIPGALN-----291
Db 157 LLSVVYQAAANLHSLVLRDVSFVGQKRGDADATINSRYNDLTRIGNYDYAVRWNTGLE 216
QY 292 -----VYFFRLVDPRRNVACIEHALEKLYIYKECCYD 324
Db 217 RVWGPDSDRWVRYNQFRRELTTLVDLVALFPNYSRRYPPIRTVSQLTREIY-----TN 270
QY 325 PYRWLTEQDGYLKG-----RQPKSP-----SITL-----DDGLVY--VVRVLVTPCK 366
Db 271 PV---LENFDGSRGSAQGIERSIRPHLMOILNSITITDAHRGYIYWSGHQIMASP-- 325
QY 367 VYFCGPEVNVNRLNRYSEDIDNLRVSVFVDEWEKLYST--DLLPK-----ASTGSGVR 420
Db 326 VGFSGPEFTF-----PLYGTMGNAAPQORIVAQLGQGV- 358
```


; PRIOR APPLICATION NUMBER: US 08/922,505
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: US 08/754,490
; PRIOR FILING DATE: 1996-11-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hybrid Delta-Endotoxin
US-09-873-873-14

Query Match 1.8%; Score 104; DB 10; Length 1177;
Best Local Similarity 18.2%; Pred. No. 1.6;
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FCGQISDDKFAVLGSTEVS-----IQFGIGLKKEFFFLSSGSADYKQLQSYENIMQVVLH 156
Db 13 YNC-LSNPEVEVLGGERIETGYTPIDISLSLTQFLLSEFVPGAGFVLGL-VDIWGI--- 67
QY 157 RPYG--QNAQFLL-----IQFGAPRIYKRL--NSCYSPFKETPDQWVRTTDFP 203
Db 68 --FGSQWDAFLVQIEQLINQRIEFAFNQAIISRLGLSNLYQIYAES-----FREWAD 120
QY 204 PSWIGLSSSLCLOFRGVRPLNPFEEFFHYAERENNTITQGTFFFSQKSAVLPNVQPP 263
Db 121 PTNPALREEMRIQF-----NDMNSALTTPALFA-----VQNVQVP 156
QY 264 -----EGISIPYKILFKISSLVQHGCIPGALN----- 291
Db 157 LLSVYVQAANLHLSVLRDVSFVGQWGFDAATINSYNDLRLIGNYTDYAVRWYNTGLE 216
QY 292 -----VYFFRLVDPRRNVACIEHALEKLYIYKECCYD 324
Db 217 RVWGPDSRDWRVYNQFRRELITLVLDIVALFPNVDSPRYPIRTYSQLTREIY-----TN 270
QY 325 PVRWLTEQYDGLK-----RQPPKSP-----SITL-----DGLVY--VRRVLVTPCK 366
Db 271 PV---LENFDGSRGSAQGIERSIRSPHLDILNSITITYDAHRGYVYWSGHQIMASP-- 325
QY 367 VYFCGPEVNVNRVLRNYSIDIDNLFVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420
Db 326 VGFSGPETF-----PLYGTMGNAAPQQRIVAQLGQGV- 358
QY 421 TNYIERILSTL-RKGFVIGDKKFEAFSSQRL--DNSVMVFASRPGOLTA----- 468
Db 359 ---YRTLSSLYRRPFNIG-----INNQLSVLDGTEFAYGTSNLPASVYRKSGTV 407
QY 469 -----NDIRAWMGDFSOIKNVAKYAAALGSGSSRETLSVLR----- 506
Db 408 DSLDEIPPNVPPRQGFGRKLSVMSFRS-----GFSNSSVSIIRAPMFSWTHRSAT 461
QY 507 -----HEIEVIPDKVH---GFSYVFSGI--GKT-----SGDFAHRVASKGLQYT 548
Db 462 PNTIDPERITQIPLVKAHTLQSGTVVVGPGFTGDIILRTSGGPFAYTIVNINGLPQ 521
QY 549 PSAFOIRGG-----YKGVGV-----DPDSSMKLSRKMSKY----- 582
Db 522 RYRARIYASTNLRIYVTVAGERIFAGQFNKMTDGTGDPPLTFQFSFYATINTAFTFMSQ 581
QY 583 ESDNKLVLGWSKYQPCYLNAR-----QLITLLSLTGKXDE 618
Db 582 SSFTVGAOT--PSSGNEYVIDREFELIPVATFEAYDLERAQKAVNALFTSINGIKTD 639
QY 619 VLEQKQKRAVDOLDAI---LHDSLKAQALEMLSPGENTNLIKAMLNGYKPDAPPEFLSM 675
Db 640 VTDYH-----IDQVSNLVDCLSEDFCLDEKRELSEKVKHAKRLSDERNLLQDPNFKGINRQ 695
QY 676 MLOTFRASKLLDLRTSRIFIPNGTMMGCLEDES--RTLEYGV-----FVQGTGAGH 727
Db 696 LDRGWGSGTDITIQRGDDVFKNENYVTLPGTDECYPTLYQKIDESKLKATRIQLRGYI 755

QY 728 EESDHLHPENSRSTNSNFILKGNVVVAKNPKLPGDIRVLKAVNVRALHHMVDVVFPQ 787
Db 756 EDSQDL-----EYILIRYN---AKH-----ETVNPVGTSLWPLSLSAQSP 791
QY 788 KGRPHPHNECS-----GSDLD 803
Db 792 IGKCGEPNRCAPHLEWNPDL 812

RESULT 13

US-09-873-873-28
; Sequence 28, Application US/09873873
; Patent No. US20020064865A1
; GENERAL INFORMATION:
; APPLICANT: Malvar, Thomas
; APPLICANT: Gilmer, Amy Jelen
; TITLE OF INVENTION: Polynucleotide Compositions Encoding Broad-Spectrum S-Endotoxin
; FILE REFERENCE: MECO:210--2
; CURRENT APPLICATION NUMBER: US/09/873,873
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: US 09/253,341
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: US 08/922,505
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: US 08/754,490
; PRIOR FILING DATE: 1996-11-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hybrid Delta-Endotoxin
US-09-873-873-28

Query Match 1.8%; Score 103; DB 10; Length 1177;

Best Local Similarity 18.2%; Pred. No. 2;
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FCGQISDDKFAVLGSTEVS-----IQFGIGLKKEFFFLSSGSADYKQLQSYENIMQVVLH 156
Db 13 YNC-LSNPEVEVLGGERIETGYTPIDISLSLTQFLLSEFVPGAGFVLGL-VDIWGI--- 67
QY 157 RPYG--QNAQFLL-----IQFGAPRIYKRL--NSCYSPFKETPDQWVRTTDFP 203
Db 68 --FGSQWDAFLVQIEQLINQRIEFAFNQAIISRLGLSNLYQIYAES-----FREWAD 120
QY 204 PSWIGLSSSLCLOFRGVRPLNPFEEFFHYAERENNTITQGTFFFSQKSAVLPNVQPP 263
Db 121 PTNPALREEMRIQF-----NDMNSALTTPALFA-----VQNVQVP 156
QY 264 -----EGISIPYKILFKISSLVQHGCIPGALN----- 291
Db 157 LLSVYVQAANLHLSVLRDVSFVGQWGFDAATINSYNDLRLIGNYTDYAVRWYNTGLE 216
QY 292 -----VYFFRLVDPRRNVACIEHALEKLYIYKECCYD 324
Db 217 RVWGPDSRDWRVYNQFRRELITLVLDIVALFPNVDSPRYPIRTYSQLTREIY-----TN 270
QY 325 PVRWLTEQYDGLK-----RQPPKSP-----SITL-----DGLVY--VRRVLVTPCK 366
Db 271 PV---LENFDGSRGSAQGIERSIRSPHLDILNSITITYDAHRGYVYWSGHQIMASP-- 325
QY 367 VYFCGPEVNVNRVLRNYSIDIDNLFVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420
Db 326 VGFSGPETF-----PLYGTMGNAAPQQRIVAQLGQGV- 358
QY 421 TNYIERILSTL-RKGFVIGDKKFEAFSSQRL--DNSVMVFASRPGOLTA----- 468
Db 359 ---YRTLSSLYRRPFNIG-----INNQLSVLDGTEFAYGTSNLPASVYRKSGTV 407


```
/ APPLICANT: Silva, Jeff
/ APPLICANT: Summers, Eric
/ TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
/ FILE REFERENCE: 109272.147
/ CURRENT APPLICATION NUMBER: US/09/801.368
/ CURRENT FILING DATE: 2001-03-07
/ PRIOR APPLICATION NUMBER: US 09/487,558
/ PRIOR FILING DATE: 2000-01-19
/ PRIOR APPLICATION NUMBER: US 60/160,587
/ PRIOR FILING DATE: 1999-10-20
/ NUMBER OF SEQ ID NOS: 440
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 370
/ LENGTH: 1331
/ TYPE: PRT
/ ORGANISM: Saccharomyces cerevisiae
US-09-801-368-370
```

```
Query Match 1.7%; Score 102; DB 10; Length 1331;
Best Local Similarity 18.8%; Pred. No. 3;
Matches 228; Conservative 170; Mismatches 413; Indels 404; Gaps 69;

Qy 44 SRAFAKVOFADNIS---ADKIITIANNLXFGS---SYLKAW-EMKTDIV-----OLRAYVD 93
Db 251 NRYHSVSVWPNLKNMKYDK2IT---NTAFTARIDVWNSWLNKFTNLTLLRRQELDDWIN 307
Qy 94 QMDGTLNFGCOISDD-----KFAVLGS---TEVSIOFGIGLKKPFFFLSSGSADYKL 143
Db 308 RPSISSNDCQDFGVPQWNCMKILAEOLMEKNE-SIFOKKIFYPUSPMW--FKL 364
Qy 144 QLSY----ENIWQVVLHRYPGNAQFLLIOLFAPRIY-----KRLNSCYSEFFKETP 192
Db 365 KLHFIVYRETLTKMKIKYPERLSRSLAAPPVLYLKEVILTRLSYARKLKNPTMMIDQMI 424
Qy 193 DDQWRTTDFPSPWGLSSSCLCFRCGV-RLP---NFEESFFHYARENNITLOTGTFP 248
Db 425 DDF-----NAFIRLSVQLYTLTKYCSNLFPDFVDFPTF-----ENTVIEAIRYLE 470
Qy 249 FV-----SOKSALVPNVQPEGISTIPYKILFKITSSLVQHG---IPGALNVYFFRLV 298
Db 471 FLNLKLIDSSKQ---NEKAPD-LLLYWDHLKNTGHYINGAETVIPNEFLKL-TLRV 524
Qy 299 DRRNRVACIHALEKLYIKECCYDPV-----RWLTEQYD--GYLKGQRPKPSPI 348
Db 525 -----HKLQ-FYLLAQONFPPTFANASEAKWLSSIFENLGAMK-RKLNRFSNI 571
Qy 349 -----TLDDG---LVVRRVLVTPCKVYFCGPEV-----NYSNR 379
Db 572 LVKAFQNSAVYQINHNALVKLKDHAHYLVYSGNTFESSGVYNAFAPELLGCNDTILR 631
Qy 380 VLNRYSIEDNFRVSVFDEWEKLYSTDLLPKASTGSGVETNIYERILSTLRKGFVIG- 438
Db 632 ILRN-----KSIGDLVPKLDIGNNL--NVYDITTKETDLNILVSK 670
Qy 439 --DKK---FEFLAFSSQL-----RDSNVW----- 458
Db 671 GEDSGIPYRVANSSDLDORHAHQSKKNFSTDPDQHLDEKNNEVFELEVALSLGA 730
Qy 459 MFASRPGLTANDIRAWMGDFSQIKNVAKYAAR---LQSGFSSRETLSVLRHEIEVIPDV 515
Db 731 LVVLYPG---EPVVWDGVPVKLPGNLNFASNEMDLGKI--GNPNTLILN----- 775
Qy 516 KVHGTSVYFSGIGKIS---GDFHRVASKCGLQYTPSAFQ-----IRY---GGYKG 561
Db 776 --QGSNYALTQIDKFNQTVGDSVSFIEKRCSLNSIESLQKINKAYKLYTYVTLNNYK 833
Qy 562 VVGVDPSDSMK-----LSLRKSMKSYESDN-----IKLDVLGWS 595
Db 834 ILG-----SFMKQCPGNELLNLSIFMGDFGRSFLKYNAFSSKRKYVIFLWKLK-MNWL 888
Qy 596 K-----YQPCYLNRLQITLLST-----LGVKDEVLEQKQKEAVOQDLAILH 636
Db 889 KFLVEECDPTQDTRFWCVLWAMDFAMQMTSGYNILALNVK-QFQELKERVSVCMSELLISH 947
```

```
Qy 637 DSLKAQAEALMSPGENTNLIKAMLCGYKPDAPFELSMMLQTFRASKLIDLTLRTRGRI 696
Db 948 FVYMGARATE---AEN-GMQQARLNDTEENIDEATLEINS--RLRLEAINTLEKTKM 1000
Qy 697 PNGRTMMGCLDESRTLEYGQVFQVFTGAGHGEFSDDLHPNNSRSTNSNFI---LKGNVV 753
Db 1001 RNPRQMGKVLDAT---DQGNKYLLSLASSLSNVS-----MRWOKRSFSGGTGQVY 1049
Qy 754 VAKNPCLHPGDIRVLKAVNVRALHVMVDCVVPQKGR-----PPHN--ECSGSGLDG 804
Db 1050 SAIN--LENGEILLAVKEIKIHDTTMMKK--IFPLKEEMTVLEMLNHPNIVQYGVVHR 1105
Qy 805 DIYFVCHWDQDMIPPRQVQPMYPPAPSIQ--LDHDTVIEVEEYFTNYIVNDSGLIANA 862
Db 1106 DKYNIF-----MEYCEGGSLASLLDHGRIDEMVTQVYTFELLEGLAYLHQS 1152
Qy 863 HVVFADREPDMAMSDPCKKLAEFLSTAVDPKTVGPAETPSQLRPKEYDPDFMCKPKTSY 922
Db 1153 GVVRHDIKPE-----NILLDF--NGI-----IKYVDF-----G 1178
Qy 923 ISERVIGKLPKVKDKAPQASSIATFTROVARRSYDADMEVDGFEDYIDEAFDYKTEYDN 982
Db 1179 TARTVVGSRTRTVRNAAVQDFGVETKSLN-----EMMGTPMYM----- 1216
Qy 983 KLGNLDYIGIKTEAEITLSSGIMKASKTFDRRKDAEALISVAVRALRKEARAFKRRNDID 1042
Db 1217 -----APETISGSVKGKLGAD---DVALGCVVLEMATGRPRWNSLDNE-- 1258
Qy 1043 DMLPKASAW---YHV 1054
Db 1259 -----WAIMYHV 1265
```

Search completed: November 6, 2002, 03:47:45
Job time : 22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:43:53 ; Search time 25 seconds
(without alignments)
4283.750 Million cell updates/sec

Title: US-09-782-874-2
Perfect score: 5864
Sequence: 1 MGKTIQVFGPPYLLSAEYWK.....RPVNLSSLRAQLSHRLVLK 1114
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*
1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5864	100.0	1114	2 T30819	RNA-directed RNA p
2	5070	86.5	1116	2 T30828	RNA-directed RNA p
3	1995.5	34.0	1133	2 T01920	probable RNA-direc
4	948	16.2	1215	2 T11660	probable RNA-direc
5	932	15.9	1780	2 T20695	hypothetical prote
6	896.5	15.3	3228	2 T21381	hypothetical prote
7	362	6.2	966	2 F84382	hypothetical prote
8	330.5	6.0	929	2 H84382	hypothetical prote
9	335.5	5.7	905	2 G84382	hypothetical prote
10	222.5	3.8	604	2 F87936	protein M01G12.12
11	222.5	3.8	604	2 T23669	hypothetical prote
12	138	2.4	1021	2 T40328	conserved hypotet
13	133.5	2.3	2364	2 T40884	cytotoxin L - Clos
14	133	2.3	2810	2 T22928	hypothetical prote
15	131	2.2	463	2 S37372	low-temperature vi
16	130.5	2.2	1091	2 F83288	hypothetical prote
17	129.5	2.2	849	2 H86430	T518.2 protein - A
18	129	2.2	1733	1 RNBY2L	DNA-directed RNA p
19	128.5	2.2	839	2 T12827	hypothetical prote
20	126.5	2.2	1610	2 T11681	hypothetical prote
21	124	2.1	551	2 AC3572	lysine-tRNA ligase
22	123.5	2.1	1038	2 AF3286	ATP-dependent DNA
23	121.5	2.1	1745	2 T48366	anthranilate phosph
24	121.5	2.1	1972	2 S68176	TGG protein - huma
25	118.5	2.0	404	2 T21251	hypothetical prote
26	118.5	2.0	503	1 DECHE	glutamate dehydrog
27	117	2.0	688	2 S05380	hypothetical prote
28	116.5	2.0	1200	2 T15921	hypothetical prote
29	116.5	2.0	2137	2 T05244	hypothetical prote

30 116 2.0 968 2 C64691 type III restricti
31 115.5 2.0 436 2 E71086 hypothetical prote
32 115 2.0 1163 2 E71481 probable swi/snf h
33 115 2.0 4589 2 T14914 dynein beta heavy
34 114.5 2.0 776 2 A11420 gamma-glutamylcyst
35 114.5 2.0 2214 1 QZBYU2 pyrimidine synthe
36 114.5 2.0 3191 2 T22945 hypothetical prote
37 114 1.9 972 2 C71826 hypothetical prote
38 114 1.9 1160 2 T39898 carbamoyl-phosphat
39 114 1.9 5107 2 T29144 partial CDS - Caen
40 113.5 1.9 633 2 T31160 maturase-related p
41 113 1.9 744 2 A81719 exodeoxyribonuclea
42 113 1.9 1748 2 S63127 probable membrane
43 112.5 1.9 1188 2 E89896 chromosome segrega
44 112.5 1.9 1816 1 S68960 laminin alpha4 ch
45 112 1.9 989 2 T03838 telomerase catalyt

ALIGNMENTS

RESULT 1

T30819
RNA-directed RNA polymerase (EC 2.7.7.48) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30819
R:Schiebel, W.; Pellissier, T.; Riedel, L.; Thalmair, S.; Schiebel, R.; Kempe, D.; Loi
Plant Cell 10, 2087-2102, 1998
A:Title: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato.
A:Reference number: Z20885; M01D:99055198; PMID:9836747
A:Accession: T30819
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-1114 <SCH>
A:Cross-references: EMBL:Y10403; NID:g4038591; PIDN:CAA71421.1; PID:g4038592
C:Genetics:
C:Note: RORP
C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase
C:Keywords: nucleotidyltransferase

Query Match 100.0%; Score 5864; DB 2; Length 1114;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVFGPPYLLSAEYWKSFLEKTYGTVCALVQKSGSRAFAKVQFADNISADK 60
DB 1 MGKTIQVFGPPYLLSAEYWKSFLEKTYGTVCALVQKSGSRAFAKVQFADNISADK 60
QY 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDMDGITLNFCCOISDDKFAVLGSTEVS 120
DB 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDMDGITLNFCCOISDDKFAVLGSTEVS 120
QY 121 IQFGIGLKFPFFLLSSGSADYKQLQSYENIHWVQLHRPYGQNAQFLLIQIFGAPRIYKRL 180
DB 121 IQFGIGLKFPFFLLSSGSADYKQLQSYENIHWVQLHRPYGQNAQFLLIQIFGAPRIYKRL 180
QY 181 ENSCYFFKRETPDDQVTRTDPFPPSWIGLSSSLCLOFRRGVRLPNFEESFFHYAERENNI 240
DB 181 ENSCYFFKRETPDDQVTRTDPFPPSWIGLSSSLCLOFRRGVRLPNFEESFFHYAERENNI 240
QY 241 TLGTGTFVFSQKSAVPNVQPEGISIPYKILFKISSLVQHGICFGPAPLVVFFRLVDP 300
DB 241 TLGTGTFVFSQKSAVPNVQPEGISIPYKILFKISSLVQHGICFGPAPLVVFFRLVDP 300
QY 301 RRRNVACIEHALEKLYIIECCYDVPVRWLTQEDGVLKGRQPPKSPSIITLDGLVYVRV 360
DB 301 RRRNVACIEHALEKLYIIECCYDVPVRWLTQEDGVLKGRQPPKSPSIITLDGLVYVRV 360
QY 361 LVTPCKVYFCGPEVNVNRLNRSYEDINFLRVSVDEWEKLYSTDLLPKASTGSGYR 420
DB 361 LVTPCKVYFCGPEVNVNRLNRSYEDINFLRVSVDEWEKLYSTDLLPKASTGSGYR 420

Db 452 IHKKLYVTPTTIRLIVEDSLSEAGNRVIRNFKDFANRFRMRVQITDEYYKQ-----KIR 503
Qy 415 TGS-GVRT-NIYERILSTLRKGVIGDKKFEFLAFSSOLRONSVMFASRPGLTANDIR 472
Db 504 GSGDGRNEKLYBRIQOOLTGYKVGNGQIYEFLEAFNSOLREHGAYFFASGSDLNKQIR 563
Qy 473 AMWGDFSOIKNVAKYAARLGQSGSSRETLSVLHRHEIEVDPVKVHGTSYVSDGIGKTS 532
Db 564 EMWGDFSEINSVSKYAARMGQCSTTKE---INRCVDISLQDDIVRNHNCFTDGVGMAS 620
Qy 533 GDFAHVASKC-GLQYTPSAFOIRYGYGKGVGVDPDSSMK-----LSLRKSMKSYSD 585
Db 621 LSVIRLSLEVKNHDMFPASQFRMGGYKGVLSLAPTLEKLEHOGNVLVPPRKSQDKRKF 680
Qy 586 NIKDLVLGWSKYQPCYINRQILITLLSTLGVKDEVLKQKQEAVDQLDALHDSLRKAEAL 645
Db 681 HSTLEVIKISFNSAHNLQMOLITLLEGLGVKTVFLELRSQLSKMNESINSKQSILML 740
Qy 646 -ELMSPGENTNILKAMLCNGYKPDAPFLSMMLQTFRASKLLDLRTSRIFIPNGRTMG 704
Db 741 RDNVDEYHSTLIITADFTQAGFLERDAFTENLLNLYEWLRIKQKQVSPKGYAILG 800
Qy 705 CLDESRTLE-----YQGVFQVGTGAGHGFEFSDLLHPFNNSRSTN--SNFILKGNVY 753
Db 801 VADEGTGLKGYDDAVLSVPEIFIQT-----DTSTSGSYSTGKLTQVIVGLCI 851
Qy 754 VAKNPLCHPGDIRVLKAVNVRALHHMVDVVFQPKGRPHNPCESGSDLDGDIYFCVWDQ 813
Db 852 VARNPSLPHGDVRVCAVRKAVRCEMLHLKNVIVFTTGDRTIPAMCSGGDLGDDEYTVINDQ 911
Qy 814 DMIPPROVQP--MEYPPAPSIQ-LDHDVTIEEVEEFTNYIVNDSLGIITANAHVFA--- 867
Db 912 RLLPKIVNPPLESPPKKSIDLEKPLIDSVKFEFVNYKYDSLGLISNAWKAHWH 971
Qy 868 DRPDMAMSDPKKLABLFSIANDVPKTPVPAEIPQSLRKEYPDPMDPKDTSYISERV 927
Db 972 DNNPEGIFGNVCELEAEMHSAVDFAKSVACQAKYHPKRYPDPMQTKTTSRSETA 1031
Qy 928 IGLFRKVKDKAPASSIAFTROVARSYDADMEVDGFE-DYIDEAFDYKTEYDNKLN 986
Db 1032 VGKIFRVAARFQRESGRPAY-NPIMNTVYDPCMKLPFRKTEVLNVAEEVKHYDNDLRS 1090
Qy 967 LMDYVIGKTBAILSGIM---KASTFTDRKDAEALSIVAVRALKEARAWFKRRNDIDD 1043
Db 1091 IMARFIDISTEYEVYATILFKDOLAKTNEYGLREEVSFQFDLLKKYTKQYLEKCALSN 1150
Qy 1044 MLPKASAWYVYTH---PTYWCYNQGLKRAH-----FISPPWCYVDOLIOIKK 1089
Db 1151 QSAFDSSEYEERINSAVAATIDVTYDORVKSNGVTTEVLSPFYLSSRLCOLSR 1206

RESULT 5
T20695
hypothetical protein F10B5.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20695; T24512
R:Sims, M.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z19310
A:Accession: T20695
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1780 <W1>
A:Cross-references: EMBL:Z48334; PIDN:CAA88315.1; GSPDB:GN00020; CESP:F10B5.7
A:Experimental source: clone F10B5
R:Burton, J.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19901
A:Accession: T24512
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1780 <W1>

A:Cross-references: EMBL:Z66500; PIDN:CAA91312.1; GSPDB:GN00020; CESP:F10B5.7
A:Experimental source: clone T05C12
C:Genetics:
A:Gene: CESP:F10B5.7
A:Map position: 2
A:Introns: 23/2; 226/3; 305/2; 343/2; 462/3; 736/1; 922/2; 993/1; 1222/2; 1463/3; 161
Query Match 15.9%; Score 932; DB 2; Length 1780;
Best Local Similarity 25.1%; Pred. No. 9.8e-5g;
Matches 322; Conservative 199; Mismatches 465; Indels 296; Gaps 44;
Qy 14 LSAEVVVKFLKTYG-YGTVCALVKQSKGSRAPAKV--OFAD-NISADKIITLANNRL 69
Db 270 LFSEFASFTGRITOMLHDQVLEVPKM---HTLTKTTPQHMNDINSAIAGNCPNSGL 326
Qy 70 YFGSYLKAWEMKTDIVQLRAYVQMDGITLNFQCIQSDDRFAVLGSTE----- 118
Db 327 FLVRGDFISQENTVCSVKLQSH-----HNADASRENSFSFKVAGSNKLYSARFEHDK 378
Qy 119 --VSTQFGIGLKKFFFLSSG--SADYKLOLSYENINQVVLHRPYGONAQFLLQLFGAP 174
Db 379 RLAVYVFGVRLAE---FADGDLHAGFLNLYNLFVRIVVDMSH-ETNSIYIQMKPP 434
Qy 175 RIYKRL-ENSCYSFFK-----ETPDQDQVRTDPP-----PSWIGLS 210
Db 435 HLWEGIPKNTIFHPSKSVLNMET-CTEWTRVLSWPGDAEGRGVGCTSEAFSQSSWIRLT 493
Qy 211 SSLCLQFRGVLPNPFESFFHYABRENNITLQGTFFVSKSALVNVQPPGISIPY 270
Db 494 MR-----KDDNDVSSTOLMDIVTKLSA-----RSKA 521
Qy 271 KILFKISSIVQHGCIIPGAL-----NYFFRLVDPRRNVAC----- 307
Db 522 KYMFGSISIRKLAPSPAFHSLGSFRANYALQALITRGSVFTDQDFDADENIPSSND 581
Qy 308 -----TERALEKLY-----IKEC-----CYDPVWLTE-- 331
Db 582 NDEDDDDVDYDKPKMELVHEPLFLKVRGMEKCSQATEETLEOLLNAFDERRIDVVT 641
Qy 332 -----QYDGYLKGQPKPSITLDDGL-----VYVRRVLVTPCKYVFCGP 372
Db 642 AFTMYQSKRIQYERLLKGES-----LDVGLAKPLPKNCVSAKVIPTSRILLMAP 694
Qy 373 EYVNSVRNRYNSEDIDNPLRVSVDEWEKLYSYDLPLKASTGSGVTNIYERILSLR 432
Db 695 EYMAVNVVYRFGPDY--ALRCVFRDDNLGLAIRDF--SINNIDHMSNIVTEGILYTLK 750
Qy 433 KGFVIGDKKFEFLAFSSOLRONSVMFASRPGL-----TANDIRAWMGDFSOIKNVA 485
Db 751 NGLQVADRVYSFLGWSNSQMRDQGCYLYAPRVNALTGEVGTGVEDIRVWMDGFRDAISVP 810
Qy 486 KYAARLGQSGSRETLSVLRRHEIEVDPVKVH-----GTSYVFSDGIG 529
Db 811 KMMSRMGQCFQAOPTVYSSVKNIHIVENIQVRLERHHWIVEPDIEGGYENKYCFSDGCG 870
Qy 530 KISGDFAHVASKGLQYTPSAFOIRYGYGKGVGVGDDP-----SSMKLSLRKSKSY-- 582
Db 871 RISIKLATHISKILQKEVPACFQVRFGKGLIILVIDTIDDIINWPVKVIFRKSOQKE 930
Qy 583 ---ESDNIKDLVGSKYQPCYLNRLITLLSTLGLVKQDEVLEQK-----QKEAVD 629
Db 931 GGGELQDEVLEVVKYKAMPSPVCLNRPFFITLDQVSEKQSASHRRITNRVHYLRELC 990
Qy 630 QLDAILHDSLRKAEALEMSPGENTNIL----KAMLCNGYKPDAPFLSMMLQTFRASKL 685
Db 991 LSNMLINENQAABELY-----NRTNLADWNAASKRAGFELSVDPLIIDLMLFSYRYNI 1044
Qy 686 LDLRTSRIFIPN--GRTMMGCLDESRTLEYQGVFQVGTGAGHGFEFSDLLHPFNNSRSTN 743
Db 1045 IHHISKAKIFLPPSLGRSMYGVVDETGLLQYQGVFIQYS-----PSIRQTS 1090
Qy 744 SNFILK-GNVVYAKNPCLHPGDIRVLKAVNVRALHHMVDVVFQPKGRPHNPCESGSD 802
A:Residues: 1-1780 <W1>

Db 1091 NRPILKTGKVLITKNPCHVPGDVRVDAVWQPALAHLVDVVVFPQHGPRPHDENMAGSDL 1150
QY 803 DGDIVFVCDQDMPQVOPVPPAPSIQDHDVTEVEVEYFTNYIVNDSGLIANA 862
Db 1151 DGEVSIINDEQMLDYNNEAMVFPSSNAEEDKPTDDVDFEFLRLVLOQDSIGRMSHA 1210
QY 863 HVFADREPDMAMSPCKKLABELFIAVDFFPKTGPVPAEIPSLRKEY-PDFMDKPKDTS 921
Db 1211 HLAYADLHG--LFHENCHAIALKCAVAVDFPKSGVPAEPLSLFQECENTPDYMMGGKPM 1268
QY 922 YTSERVIGKLFKRVKDKAPQASSIAFTTRDVARRSYD-----ADMEV-DGPEYIDEAFD 975
Db 1269 YSTRUNGQLHKKAR-KVEEVLBEETRGVSFEYDKLICPEDVDVDFGNEIKVLQILT 1327
QY 976 YKTEYDNKLGNIYMGIGKTEAEILSGGIMKASKTFDRKDK-----AEASIVAVRALR 1028
Db 1328 LRDEYVDRMQLLDEYIEDEASVSGHAASIKRLAGMERDDYSFVHTDKVVELRYEKLY 1387
QY 1029 KEARAWF-----KRRNDDMLPKASAWHYHTY-HPTY---WCYNOG 1067
Db 1388 AVRAKFFEEFGGEEINENIDGKNTLCKTKAMHEKIRQWFEVAVVQPRINKAGRCIGQ- 1446
QY 1068 LKRAHFISPPWCYVDOLIQIKK 1089
Db 1447 -----SLPWAWDALCDLRR 1461

RESULT 6

T21381

hypothetical protein F26A3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21381

R:McMurray, A.

submitted to the EMBL Data Library, August 1996

A:Reference number: Z19415

A:Accession: T21381

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-3228 <WIL>

A:Cross-references: EMBL:Z78419; PIDN:CAB01701.1; GSPDB:GNO0019; CESP:F26A3.3

A:Experimental source: clone F26A3

C:Genetics:

A:Gene: CESP:F26A3.3

A:Map position: 1

A:Introns: 36/3; 91/3; 216/3; 263/3; 308/3; 702/3; 828/3; 866/1; 1013/3; 1171/3; 1231/3;

; 2443/1; 2595/3; 2663/1; 2753/3; 2813/3; 2859/1; 3052/3; 3136/1; 3176/1

Query Match 15.3%; Score 896.5; DB 2; Length 3228;

Best Local Similarity 24.8%; Pred. No. 6.7e-52;

Matches 314; Conservative 210; Mismatches 396; Indels 345; Gaps 50;

QY 69 LVFGS-----SYLKAWEM-----KTDI---VQLRAYVDQMDGITLNFQC-QISDDKFA 112

Db 1732 IFPGNLOGGTYNHNWVSRENISRDMLKHAEFEFDKTDMTVQFCEEKKQFE 1791

QY 113 VLGSTEVSIQFGLKFFFLSSGADYKQLSYENIWQVYLH-RPYGQNAQFLLIQLF 171

Db 1792 DSRKQVRV-----NYQLTIRDSIRRIIVDPVEGCNT-CVHFEVN 1832

QY 172 GAPRIYK-LENSCYSEFFKETDQDQVWRTDFFPSH-----IGLSSSLCLO 216

Db 1833 CPLLIRKVIDNDKSFHKPF-----YEROKRFCDWRNGNVNHNQPDAAIADSPFFTIE 1888

QY 217 FRGVGLPNFESFFHYAERENITITQGTFFVSKSALVPNVQPPGESIPY----- 270

Db 1889 FHKESI-----TKEMRYVLSLRSTK-----VLIEFANLPSIDVPMGSHYPNWNK 1937

QY 271 -----KILFKISSLVQHGCIFGPAL- 290

Db 1938 KSPTDSNAPIFREFLKEIFPPKYEIVDDKLIDVNEERKFSITYLIECLLSRGAIVKDQIL 1997

QY 291 -----NVYFER-----LVDPRRNVACIEHALEKLYIYKECY 323

RESULT 7

F84582

Hypothetical protein At2g19910 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

Db 1998 LNEQHWKNFLEIITWYRNNDQNLCEAALEDVLHLIDGKR-IGSILKCLDKICOKRE-VM 2055
QY 324 DPVRLWTEQ--YDGYLKGROPPKSPSITLDDGLVYVRRVLVTPCKVYFCGPEVNVSRVL 381
Db 2056 KLVNGLTEKESIEGYR-----VRKVFTPTFVYIAPETIMGNRVL 2097
QY 382 RNYSEDIDNFRVSVFDEWEKLYSTDLLPKASTGSGVRNIYERILST-----LRKGFV 436
Db 2098 RFPKDGTRVLRVTRFDNNKKM-----RSNVTGKLLDRTANKYLEHVR 2142
QY 437 IGDKKFEFLAFSSSOLRDNVW-----MEASRPGLT-----ANDIRAWMGD 477
Db 2143 IANREYFLGCSNQRMDNGAIFWMRTDQKDRFYCNPTASNINPKKIDEVRFQLGR 2202
QY 478 FSOIKNAVYAAARLQSGFSGSRETLSVLRIHE-----IEVPDVKVHGTSYVFSOGIKIS 532
Db 2203 FSEIENVKMLARLQCGFTQSRLTGVLGRDDYCTVDLTGCRATINGSEYTFSDGVGMS 2262
QY 533 GDFAHVASKGLOIY---TPSAFOIRYGGYKGVGVDP-----DSSMKLS-- 574
Db 2263 YQFAQESQ---AMQFGKAVPSCFOIRFNGKGVIAIEPFLDEIRKVALNVGVTSKMAKC 2320
QY 575 -LRKSMKYESDNIKLDVLGHSKYQPCY--LNRQLITLITLSTGVKDEVLEQKQEAVDQ 630
Db 2321 LFRPSQIKFOAKISGDOJEMVKFSSAVLVALNKPFINIL-----DQVSEMQSILDCR 2374
QY 631 LDAILHDSLKAQBALELMSPGENT---NILKAM-----LNCGYKPDABEFLSMWL 677
Db 2375 ITSRIEELMDRQILSFAKQMNTEETFCRNKLEFFRRIDIDNLRITMGTUJSSSEFFRSLI 2434
QY 678 QT---FRASKLLDLRTSRIFPN--GRTMGCLDESRTELEYGVQVFTGAGHGEFSD 732
Db 2435 KASIKFSITKOL---CKEQIPISELGSMGVDEYGRLOYGQIFVQYT---KNYKKK 2487
QY 733 LHPFNNSRSTNSNFIKGNVYVAKNCPHGDIVLKAHVNRALHMHVDCVFFQKGRP 792
Db 2488 LPPRDSNNKHVSGSEITVGTLLTKNPCIVPDGVRIEAVDIPELHMKDGVVYVFPQHGRP 2547
QY 793 HPNCSGSLDGDIVFYVCWDDMIPPROVQPMYPPAPSIQDHDVTEIEV--BEYFTNY 850
Db 2548 HPDEMAGSDLDGDYSYVWDQELLERNEEFPDF-AVEKIKVYDREKLDVLMREFVY 2606
QY 851 TVNDSLGIITAHAVYFADREPDMAMSDCKKLABELFIAVDFFPKTGPVPAE----- 900
Db 2607 LKLDVSGVGISNHLHNSDQY--LNSRVCMDLAKKNCQAVDTFKSGOPDPLETKWRADP 2664
QY 901 -----IPSQLPKPEYDPF-MDKPKTSYISERVIGKLF-----KVKDKAPQASSI 945
Db 2655 VTFEVIPPE-NPERIPDFHMGNERSPMYSPRLCGKLFREFQAIDNVIKISEERDEQYNI 2723
QY 946 ATFTDRVARRSYDADMEYDGFEDYIDEAFDYKTEYDNKLGNIYGYIKTEAEILSGGIM 1005
Db 2724 -----ELDETIFVTGFERYMESAQKQLSSYNGQLRSIMENYIGIRSEIGMSGIV 2773
QY 1006 K-ASKTFDRKD-----AEASIVAVRALRKEARWF----- 1035
Db 2774 EMNRISDKDDDDSYFNNTNMIEETKMTSLVCKREFIEFEFGFTVKCTLLPNAYDNGN 2833
QY 1036 ---KRRNDDID-DMLPKASAWHYHTYHPTYWGCYN--QGLKRAHFISPPWCYVDOLIQIKK 1089
Db 2834 CLNYRCEPDQEVKKA VAWYR-----ACYECAQSTREVRKLSFAWIAVDYIAKVK 2885
QY 1090 DKARN 1094
Db 2886 TNVLN 2890

QY 688 LR-----TRSRIFPNGRMTMMGCLDESTRLEYGVGFVQFTGAGHGFSDDLHPFNNSRSTN 743
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 592 LEKDDLRAGKLPIDESYYLNGTVDP- - - - -GELKD- - - - -623

QY 744 SNFLIKGNVVANNPCLPHPGDRIVLKANVRALHHVD- - -CVVPFQPKRKPRHNPCECS 799
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 624 -----EVCILNPGPLHFGBDIIHLKATYYKSLEQYVGNSKYGVFFPQKPGRSLGDEIAG 676
| : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 800 SLDLGDIYFYCWQQDMT- - - - -pprov- - - - -QMIEYPFPAPSIQLDHHDVTIEE 842
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 577 GDFCDMYITSRNPCKLLEHYKPSPEWPWSSSPRSKIYTGROPSLSPE- --QLE- ---EE 728

QY 843 VVEEYF--TNIVYNDLSGLIIIANAHVYFADP- - - - -EPDMAMSDPCKKLAELFSIAVD 891
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 729 LFKIFLTGTGPSVSVIGQAADSWLAIMDRFLTJTDGENVKREKAEMKKMLKLTIDYIDAID 788
| : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 892 FPKTGVAETPSQLRPXEYPDFMDKDKTKTSVISERVIGLKFRKV- ---KDKAPQA 942
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 789 APTGTENVLPLOVKVDLFPHYMER-NKT-FKSTSILGLIFTDVDFHNAEITTPSA 842

RESULT 10

F87936

protein M01G12.12 [imported] - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001

C/Accession: F87936

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A>Title: Genome sequence of the nematode C. elegans: a platform for invest

A/Reference number: A75000; PMID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/_

A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2

A/Acession: F87936

A>Status: preliminary

A/Molecule type: DNA

A/Residues: 1-604 <STO>

A/Cross-references: GB:chr_I; PIN:CAB04618.1; PID:g3878658; GSFD:B:GN00019

A/Note: cDNA EST yk406n6.5 comes from this gene

C/Genetics:

[illegible]

QY 983 KL--GNLMDYYGKTBABILSGIMKASKTFDRR-----KDAEASVAVRALRKAR 1032
Db 483 KVKEGVSSKF-IAVELDSLGAFAFANAKQYELRREALIKETKTAFAASKALKSTQRKIE 541
QY 1033 ANFKRRN--DIDDMLPKASAWYHVTHPYWGYCINOG 1067
Db 542 QDLKRSTTADTQRIILLGRKTFEKEFH---WFISSEG 575
RESULT 13
I40884
cytotoxin L - Clostridium sordellii
C:Species: Clostridium sordellii
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40884
R:Green, G.A.; Schue, V.; Montell, H.
Gene 161, 57-61, 1995
A:Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium sordellii
A:Reference number: I40884; MUID:95369733; PMID:7642137
A:Accession: I40884
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2364 <RES>
A:Cross-references: EMBL:X82638; NID:g1000694; PIDN:CAA57959.1; PID:g1000695
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin
Query Match 2.3%; Score 133.5; DB 2; Length 2364;
Best Local Similarity 18.9%; Pred. No. 4.7; Mismatches 365; Gaps 57;
Matches 208; Conservative 142;
QY 47 FARVQPADN--ISADKIITIANRLYFGSSYLKAWEMKTDIVQLRAYVDQ-MDGTILNFG 103
Db 189 YKKSQIEPNEPFIIDNII---KTYLSNEYSK-----DLEALNKYIEESLNKITANNNG 237
QY 104 CQI-----SDDKFAVLGSTEVIQIGLKKFFFLSSGSADYKQLSYENIWQVVLHR 157
Db 238 NDIRNLEKFADEDLVRLYNOELVERNLAAASDILRSMLEKDGYYLVDVILPGI---- 293
QY 158 PYGONAQFLLIQFGAPRIYKRENSCYSEFFKETPDQDQVTRTDFPPSWIGLSSSLCLOF 217
Db 294 -----QPDLEKSINK-----PDS---ITNT-----SWEMIKLEAIMKY 323
QY 218 RRGVRLPNFEESPFHYAERENNITLGTGTFVFSQKSAALVPNVQPEGISI---PKILF 274
Db 324 KE--YIPGYSKFNEDLDEE---VQSFESALSRSKSEIPLPLDDIKVSPLEVKIAF 377
QY 275 KISLVHGHGICPGPALNVFFRLVDRRRNVACIEHALEKLYIKCECCYDPVRLWTEQYD 334
Db 378 ANNSVINQALI-----SLKDSY-----CSDLV---INQIK 404
QY 335 GYLKGROPPKSPSITLDDGLVYVRRVLVTPCKYVFGPEVNVSNRVLNRYSEDIDNFLRV 394
Db 405 NRYKILNDLNPST--NEGTFD-----NITMKI-FSDKLASINEDNMFMKIKITNLYKV 456
QY 395 SFVDEWEKLYSTDLPLPKASTGGVTRNIYERILSTLRKGFVIGD-----KXF 442
Db 457 GFAPD---VRSTINL---SGFGVYTGAYQDILL-----MFKDNSTNIHLEPELRNF 501
QY 443 EFLAFSSQGLRD---NSVMWFASRPGLTANDIRAWMGDFSOIKNVAKYAARLQSGFSSR 499
Db 502 EFPKTKISQTEQEITSLNSF-----NQARA-KSQFEYKK-GYFEGALGEDDNLDF 551
QY 500 ETLVLRLHEI---EVIPDVKVHTSVYVSDGIGKISGDFAFHRVASKGL-----QYTPSAF 552
Db 552 AQNTVLDKVYSKILSMKTRNKEYIHY--IVQLQGDKISYEAS-CNLFSDKDPYSSILY 608
QY 553 QIRYGGVKGVGVD-----POSSMKLSLRKSMKY-----ESDNIKLDVLGSKYQ 598
Db 609 Q-----KNIEGSTAYYYVADAIEI-----KEIDKTRIPYQSNRKNIKLTFIGHGK-- 655
QY 599 PCYLNROLITLLSLGVKDEVLEQKQKAEVQDLDAILHDSLKAAQEALELMSGPENTNLIK 658

Db 656 -----SENTDTFANLDV---DSLSSE---IETI-----LNLAK 683
QY 659 A-----MLNC---GYXPDAEPFLSMMLQTFRASKLLDLRTSRFIPNGRTMGMC 705
Db 684 ADISPKYIEINLLGNCMFYSISAE-----ETYPGKLLLIKIDRVSELMP-----I 730
QY 706 LDESRTLEYGVFOFTGAGHGEFSDLLHPFN-----SRSTNSNFILKGNVVVAK 756
Db 731 SODSITVSANQYEVNINEEGKREILDHSGKWKINKEESTIKDISKEYISFPNPKENKILVK 790
QY 757 NPCLHP-----GDIRVLKAVNVRLAHMHVDCVFPQKGRPHPHNCCSGDLDGDIYFCW 811
Db 791 SKYLHELSTLLQELR-----NNANSSDIDLE----- 816
QY 812 DODMTPPPQVQPMFPAPSPAPSTQLDHVDVIEVEEYFTNYIVNDSIGIITANAHVFA---- 867
Db 817 -----KKVMLTECEINVANSDIROIVEGRIEE--AKNLTSDSYINKNEFKLIESISD 867
QY 868 ---DREPDMASD-----PCKLAELFSIAVDFPKTG-----VPAEIPSQLR---PK 908
Db 868 SLYDLKHQGLDDSHFISFEDISKTENGFRIRFINKETGNSIFTETEKEIFSEYATHISK 927
QY 909 EYDPDMKPDKTSYISERVIGLFRKVKDKAPQASSIATFTRDVARRSYDADMEVDGFED 968
Db 928 ETSNIKDT-----IFDNVNGKLVKKV-----NLDAAHEVNTLNS 961
QY 969 --YIDEAFDYKTEYDNKLGNLMDYGIKTEAELISGGIMKASKTFDRRKDAEASVAVRA 1026
Db 962 AFFIOSLEYNTKES--LSNLSVAMKVQVYQALFSTGL---NTTIDASKVVVELYSTAL-- 1015
QY 1027 LKKEARAWFKRRNDIDDMPLPKAS 1049
Db 1016 -----DETIDLLPTLS 1026
RESULT 14
T22298
hypothetical protein F46C3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T22298
R:Cottage, A.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19543
A:Accession: T22298
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2810 <WIL>
A:Cross-references: EMBL:Z66563; PIDN:CAA91469.1; GSPDB:GMO00028; CESP:F46C3.3
A:Experimental source: Clone F46C3
C:Genetics:
A:Gene: CESP:F46C3.3
A:Map position: X
A:Introns: 67/2; 172/3; 204/2; 310/3; 366/2; 431/2; 684/3; 739/3; 786/3; 823/2; 880/1958/3; 1999/1; 2078/1; 2117/3; 2159/2; 2220/1; 2269/3; 2306/1; 2399/2; 2444/1; 2488/
Query Match 2.3%; Score 133; DB 2; Length 2810;
Best Local Similarity 17.7%; Pred. No. 6.7; Mismatches 344; Gaps 46;
Matches 169; Conservative 141;
QY 252 QKSALVPNVQPP-----EGISIPYKILFKISSLVQHGCIPGPAFN-----VYF 294
Db 1998 RKEPLVP--QPPRPVVAEKPAVKFVAPKWLTIIR-----KEMFYPGVNLNDIQDQV 2050
QY 295 FRLVDPRRNVACIEHALEKLYIKCECCYDPVRLWTEQYDGLVKGRQP-----KSPSITL 350
Db 2051 AQIVEDCKKSY-----YRIREQDKQVETVLRQYE-----IPPSDNNOSNIHP 2095
QY 351 DGLVYVVRVLVTPCKVYF-----CGPEVNVSN-----RVLNRYSEDIDNFLR 393
Db 2096 DVKVAVIELARLWP--LIFNQYEVVEKRPDESVTIFAISEGILFIVHTPHDLENPLK 2153
QY 394 VS-----FVDEWEKLYSTDLL-----PKASTGSGVTRNIYERILSTLRK---GFVI 437

Db 2154 IDPFPEETIADVSEANDILSVHVRHEDEENAYSAVRIKTNQAPQIKTKTIDRCLSGGVV 2213
QY 438 GDKFEFLAFSSQLRDSNVMFASRPGLTANDIRAMWDFSQKNVAKYAARLQSGSFGS 497
Db 2214 PKRKE-----VLARTGREPPVGNWL--YGKIEN--RFGFLLAQYVDS 2252
QY 498 S-RETLVLRHEIEVIPDKYHGTSYVFSGIGKISGDFAHVASKGCLQYTPSAFQIRY 556
Db 2253 TGDNVPPIRHETSDRDERVR-----FFDDEV----- 2280
QY 557 GYGKGVGVDPSSMKLSLRKSMKY-----ESDNKLDVLGWSKYOPCYLNRLQI----- 607
Db 2281 -----PFSSERYTMIDFATKRYFKPKDKKKQETWAMEDIS-----QIVREFSEKP 2324
QY 608 ---TLLSTLGKDEYLEQKQKAEVDQDAIL-----HDSLKQAEALELSPGENTNILKAM 660
Db 2325 ISQSLLADLG-----NEESYAVETFAIMKFMGDEPLKKSES-----TDVFKVL 2371
QY 661 LNCGYKPA--EPFLSMMLQT-----FRASKLLDLRTSRIFIPNGRMTMGCLD 707
Db 2372 LICHQPTLRDEVCYQLIKQITSNISQKPNALSARWLLTIT--AVFPSSLTK----- 2424
QY 708 ESRILEYGVQVFTGAGHGFSDDLHPFNNSRSTNSNF-----I 747
Db 2425 -----PYVLQYLGDNADENQRPFH--GTARICQTNMIQTFRYGRKVLNNALEVQOI 2474
QY 748 LKGNVY-----VAK-----NPLCHPGDI-----RVLKAVNVRALHH 778
Db 2475 TVGSLVTLFYFCBSTGCGQLRQCAFYSKDHNVQSQTLRTITVAEMIQLCNLLNVRSLHE 2534
QY 779 MYDCVVFYQKGRPHNCSGSLDGDIFYVCWQDMMTPPRQ-----VQPMYPPAP 830
Db 2535 QQEFSLCYTVGDKHLNCKNDNYLMDI--ITESHKLLPFQFYLVKRTVWVHPLRYDNAA 2592
QY 831 STQLDHDVTIEVEEYFTNYIVN--DSLGLIANAHVVFADREPDMAMSDPCKLAELFESIA 889
Db 2593 YDSMFD---QVDDYLRGSLISTNSLGLTAA-----TTEIILKALYFLLL 2638
QY 890 VDFPFTGVPAPETPSLRPEY--PDFMDKPKDTSYISE--RVIGLFRKVKDKAPQASSI 945
Db 2639 PNPKE--GLNAKTLPOIVPKSVIEPKRHOEMVTRISQLKMFGRMRPAEAKSHFLELL 2697
QY 946 AIF-----TROVAR--RSYDADMEVDGF 966
Db 2698 STWPTFGVLHYRLKSVENGHQLPEVILTINKSGIQLLOPKSKEVFKERNYDQIVSVESI 2757
QY 967 EDYI-----DEAPDYKTEYDNKLGMLMDYYGIKTEAEILSGGIMKASKT 1010
Db 2758 RKTAKIVRLVINTMQGETLIDIKDEADEISHLIGQY-----MFVTGGAEERGST 2808

RESULT 15

S37972
low-temperature viability protein LTV1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKL143w; protein YKL2
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 06-Feb-1998
R:Ramezani Rad, M.; Xu, G.; Kirchrath, L.; Fritz, C.; Keuchel, H.; Hollenberg, C.P.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37953
A:Accession: S37972
A:Molecule type: DNA
A:Residues: 1-463 <RAM>
A:Cross-references: EMBL:228143; NID:9486244; PID:9486245; MIPS:YKL143w
R:Abraham, P.R.
submitted to the EMBL Data Library, August 1993
A:Reference number: S38996
A:Accession: S38997
A:Molecule type: DNA
A:Residues: 1-134, /ATCVSKRDGKA',146-463 <ABR>

A:Cross-references: EMBL:225464; NID:9396434; PID:9396436
R:Abraham, P.R.; Mulder, A.; van't Riet, J.; Plant, R.J.; Raue, H.A.
Yeast 8, 227-238, 1992
A:Title: Molecular cloning and physical analysis of an 8.2 kb segment of chromosome 2
A:Reference number: S22274; MUID:92245761; PMID:1574929
A:Accession: S22275
A:Molecule type: DNA
A:Residues: 1-134, /ATCVSKRDGKA',146-313, /LSKTKARRGGRKRAVAKRA',334-363, /SSMVTKL' <AB>
A:Cross-references: GB:S97962
C:Genetics:
A:Gene: SGD:LTV1
A:Cross-references: SGD:S0001626; MIPS:YKL143w
A:Map position: 11L

Query Match 2.2%; Score 131; DB 2; Length 463;

Best local Similarity 18.0%; Pred. No. 0.52; Indels 206; Gaps 24;
Matches 96; Conservative 75; Mismatches 155;

QY 530 KISGDFAH-RVASKGCLQYTPSAFQIRYGYGVGVGVPDSSMKLSLRKSMYSKDNIX 588
Db 59 KPGRRRAHVGAEALYGINFDDSEYD--YTQHLKPIGLDPENSIFIAKSGNEQKVEKKNIE 116
QY 589 LDVLGWSKYQPCYLNRLITLLSTLGVKDEVLQKQKQKAEVDQDAILHDSLKQAEALELM 648
Db 117 -DLFIEPKYR-----RDEI-----EKDDALPVPQGRMAKPE---YLL 149
QY 649 SPGENTNILKAMLCNGYKPAEPFLSMMLQTFRASKLLDLRTSRIFIPNGRMTMGCLDE 708
Db 150 HQQDTTDEIR-----GFKPDMNPALREVLEAL-----EDEAVVNDVVVEDISK 194
QY 709 SRLLE---YQO-----VFVQFTGAGHG-----EFSDDLHPFNNSRSTNSNFIKGNVVA 755
Db 195 KTLQSGDNYGEEKEDDIFAQLLSGEAKDEDEFEDE----- 231
QY 756 KNPCLHPGDIRVLKAVNVRALHNMVDCVFPQKGRPHNECSGSLDGDIFYVCWQDQM 815
Db 232 -----FDEWDIDN 239
QY 816 I-----PPROPVQPMYPPAPSIQLDHVDVTIEVEEYFTNYIVNDSLGLIANAHVVFADREP 871
Db 240 VENFEDENYVVKEM-----AQFDNIENLEDLEN-----IDYQA 271
QY 872 DMAMSDPCKKLAELEFSTAVDFPKTGVPAETPSOLRP--KEYPDFMDKPKDTSYISERVIG 929
Db 272 DVRRFQKDNILEKHNSDDFSNAGLDSVNPSEEDVGLGELPSIQDK-SKT-----G 322
QY 930 KLFRKVKDKAPQASSIATFTRD---VARRSY-----DADMEVDGFEDYIDE----- 972
Db 323 KKKRKSQKKGAMSDVSGFSMSSSAIARTETMTVLDQYDQIINGVENYEELEDEBEQN 382
QY 973 --AFDY---KTEYDNKLGMLMDYYGIKTEAEILSGG---IMKASKTFDRRKA 1017
Db 383 YQPDFMSAERSDFESMLDDFLDNY-----ELESGRKRLAKKDKTEIRLEKA 428

Search completed: November 6, 2002, 03:46:51

Job time : 40 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:40:53 ; Search time 15 seconds

(without alignments)
3080.309 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVFGPYLLSAEVYK.....RPVLNLSLRAQLSHRLVLK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	131	2.2	463	1	LTV1_YEAST	P34078 saccharomyc
2	129	2.2	1733	1	RPBL_YEAST	P04050 saccharomyc
3	121.5	2.1	2032	1	CTOG_HUMAN	Q14008 homo sapien
4	118.5	2.0	503	1	DHE3_CHICK	P00368 gallus gall
5	117	2.0	688	1	YEQ7_YEAST	P40050 saccharomyc
6	116.5	2.0	1200	1	MOG5_CABEL	Q09530 caenorhabdi
7	116	2.0	950	1	ORP1_HUMAN	Q9BXW6 homo sapien
8	114.5	2.0	2214	1	PVRL_YEAST	P07259 saccharomyc
9	113.5	1.9	878	1	SWA_THETN	Q8AH4 thermoanaer
10	113	1.9	787	1	XPKA_LACPE	Q337f6 lactobacilli
11	113	1.9	1748	1	YNR2_YEAST	P33886 saccharomyc
12	112.5	1.9	988	1	TERT_SCHPO	Q13339 schizosacch
13	111.5	1.9	891	1	DPO1_HELPY	P56105 helicobacte
14	111.5	1.9	1816	1	LMA4_HUMAN	Q16363 homo sapien
15	111	1.9	648	1	RAD4_SCHPO	P32372 schizosacch
16	111	1.9	1014	1	ODOL_YEAST	P20967 saccharomyc
17	109.5	1.9	1649	1	NSG2_ACHCY	P34127 achromobact
18	109.5	1.9	1579	1	SRK2_YEAST	P33599 saccharomyc
19	109.5	1.9	4085	1	RRPA_CVH22	Q05002 human coron
20	109	1.9	501	1	DHE3_BOVIN	P00366 bos taurus
21	109	1.9	909	1	ODOL_BUCAI	P37388 buchera ap
22	109	1.9	1170	1	SMC2_YEAST	P38989 saccharomyc
23	109	1.9	1201	1	RPOC_LISIN	P7879 lysteria in
24	109	1.9	1887	1	FAS2_YEAST	P19097 s fatty aci
25	108.5	1.9	1188	1	YEC5_YEAST	P39991 saccharomyc
26	108.5	1.9	1835	1	DURL_YEAST	P32528 saccharomyc
27	108	1.8	734	1	METE_THEMA	Q9X112 thermotoga
28	108	1.8	1607	1	M3K4_HUMAN	Q9Y6r4 homo sapien
29	108	1.8	1916	1	RIF1_YEAST	P39539 saccharomyc
30	108	1.8	3061	1	POLG_PVYHU	Q02963 p genome po
31	107.5	1.8	755	1	P100_HSV7J	P25159 human herpe
32	107.5	1.8	1630	1	MSF1_PLAFW	P04932 plasmodium
33	107.5	1.8	1639	1	MSF1_PLAFW	P04933 plasmodium

34 107.5 1.8 1886 1 POL_COVMV P19199 commelina y
35 107 1.8 1357 1 RPOB_PSEAE Q51561 pseudomonas
36 107 1.8 2493 1 YBA4_YEAST P35194 saccharomyc
37 106 1.8 770 1 TOPI_THEAC Q9HMO8 thermoplas
38 106 1.8 1201 1 RPOC_LISMO P08Y96 listeria mo
39 106 1.8 1234 1 CFAH_MOUSE P06909 mus musculu
40 106 1.8 1290 1 BXB_CLOBO P10844 clostridium
41 106 1.8 1468 1 CHD1_YEAST P32657 saccharomyc
42 105.5 1.8 495 1 SYC_AQUAE P67163 aquifex aeo
43 105.5 1.8 813 1 VGLH_HSVMD O05783 mycobacteri
44 105 1.8 456 1 FPRH_MYCTU
45 105 1.8 1240 1 YNJ1_YEAST P53935 saccharomyc

ALIGNMENTS

RESULT 1
LTV1_YEAST STANDARD; PRT; 463 AA.
ID LTV1_YEAST
AC P34078;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Low-temperature viability protein LTV1.
GN LTV1 OR YK1143W OR YK12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92245761; PubMed=1574929;
RA Abraham P.R., Mulder A., Van't Riet J., Planta R.J., Raue H.A.;
RT "Molecular cloning and physical analysis of an 8.2 kb segment of
RT chromosome XI of Saccharomyces cerevisiae reveals five tightly linked
RT genes.";
RL Yeast 8:227-238(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Rad M.R., Xu G., Kirchhuth L., Fritz C., Keuchel H., Hollenberg C.P.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO S.POMBE SPAC3F10.17.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z25464; CAA80955.1; -;
DR EMBL; Z28143; CAA81984.1; -;
DR PIR; S37972; S37972.
DR SGD; S0001626; LTV1.
FT CONFLICT 135 145 LPVFORGMKP -> ATCVSKRDGKA (IN REF. 1).
SQ SEQUENCE 463 AA; 53390 MW; F432C239B3B0A68A CRC64;

Query Match 2.2%; Score 131; DB 1; Length 463;
Best Local Similarity 18.0%; Pred. No. 0.19;
Matches 96; Conservative 75; Mismatches 155; Indels 206; Gaps 24;
QY 530 KISGDEFAH-RVASCGQYTPSAFOIRYGVKGVGVDPDSSMKLSLRKSMKSYEDSNIK 588
DB 59 KPGRRRAHVEALYGINFDSEYD--YTQHLKPIGLDPENSIFIAKSGNEQKVEKNIE 116
QY 589 LDVLGWSKYQPCYLNRLQTLTLSTLGVLEQKQEAVDQLDLHLSLKQAEALELM 648
DB 117 -DLFIEPKYR-----RDEI-----EKDDALPVFORGMKP----YLL 149
QY 649 SPGENTNLLKAMLCNGYKPDAPFELSNMQLTFRASKLLDLRLTRSRIFIPNGRTMAGCLDE 708

Db 150 HQDQTTDIR-----GFKPMNPALREVLEAL-----EDEAYVNDVVVEDISK 194
 Qy 709 SRTLE---YGO-----VFQFTGAGHG-----EFSDDLHPFNNSRSTNSFILKGNVVA 755
 Db 195 KTOGDNNGEEREKEDIFAQLLSGEAKDEFEDE----- 231
 Qy 756 KNPCLHPGDIRVLKAVNRALHHMVCVFPQKGRPHNCEGSLDGLDIYFVNDQDM 815
 Db 232 -----FDEWDIDN 239
 Qy 816 I---PPRQVPMYPPAPSTQLDHDVTIEVEYFTNYIVNDSLGLIAHNVVVFADREP 871
 Db 240 VENFEDENYKEM-----AQFDNIENLEDLEN-----IDYQA 271
 Qy 872 DMAKSDPCKKLAEILFSAVDFPKTGPVPAIPSQLRP---KEYPDMKPKDKTSYISERVIG 929
 Db 272 DVRRFQKDNSILEKHSNDEFSNAGLDSVNPSEEDVLGELPSIQDR-SKT-----G 322
 Qy 930 KLFKVKDKAPQASSIATETRD-----VARRSV-----DADMEVDGPFEDYIDE----- 972
 Db 323 KKKRSQKKGAMSDVSGFSMSSSAIARTETMTVLDQYDQIINGYENYEELEEDBEQN 382
 Qy 973 --AFDY---KTEYDNKLGMLMDYTGKITEAELSGG---IMKASKTFDRKDA 1017
 Db 383 YQPFDMGAERSDFESMLDDFLDNY-----ELESGRKLAKKDEIERLKEA 428

RESULT 2
 APBI_YEAST STANDARD; PRT: 1733 AA.
 ID RPBI_YEAST 04050; Q12364; Q92315;
 DC P04050; Q12364; Q92315;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6) (B220).
 GN RPBI OR RPO21 OR RPB220 OR SUA8 OR YDL140C OR D2150.
 OS Saccharomyces cerevisiae (baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID:4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A364A;
 RA MEDLINE=85282617; PubMed=3896517;
 RX Allison L.A., Moyle M., Shales M., Ingles C.J.;
 RT "Extensive homology among the largest subunits of eukaryotic and
 RT prokaryotic RNA polymerases";
 RL Cell 42:599-610(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288c / FY1679;
 RX MEDLINE=97127826; PubMed=8972577;
 RA Woelfl S., Haneman V., Saluz H.P.;
 RT "Analysis of a 26,756 bp segment from the left arm of yeast
 RT chromosome IV";
 RL Yeast 12:1549-1554(1996).
 RN [3]
 RP SEQUENCE OF 1669-1733 FROM N.A.
 RC STRAIN-S288c;
 RX MEDLINE=95377607; PubMed=7649444;
 RA Croan J.E. Jr., Wallace J.C.;
 RT "The gene encoding the biotin-apolipoprotein ligase of Saccharomyces
 RT cerevisiae";
 RL FEMS Microbiol. Lett. 130:221-230(1995).
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(n).
 CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
 CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.

CC THE PHOSPHORYLATION ACTIVATES POL2.
 CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND TRNA GENES.
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X03128; CA226904.1; -
 CC EMBL: X96876; CA655619.1; -
 CC EMBL: Z74188; CA98713.1; -
 CC EMBL: U27182; AAC49058.1; -
 CC PIR: A00692; RNBX2L.
 CC SGD: S0002299; RPO21.
 CC InterPro: IPR000684; RNA_polIII_repeat.
 CC InterPro: IPR000722; RNA_pol_A.
 CC InterPro: IPR002879; RNA_pol_A2.
 CC Pfam: PF00623; RNA_pol_A; 1.
 CC Pfam: PF01854; RNA_pol_A2; 1.
 CC KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
 CC DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
 CC FT ZN-FING 67 83
 CC FT DOMAIN 1544 1719
 CC FT VARIANT 1653 1659
 CC FT CONFLICT 1514 1514 A -> V (IN REF. 1).
 CC FT CONFLICT 1524 1524 G -> A (IN REF. 1).
 CC FT CONFLICT 1601 1601 T -> M (IN REF. 1).
 CC SQ SEQUENCE 1733 AA; 191610 MW; A45C1360FF99F968 CRC64;
 CC
 CC Query Match 2.2%; Score 129; DB 1; Length 1733;
 CC Best Local Similarity 17.1%; Pred. No. 1.9; Indels 376; Gaps 48;
 CC Matches 204; Conservative 167; Mismatches 167; Gaps 48;
 CC
 CC QY 57 SADKIITLANNR-----LYFGSSYLKAWEMKTDIVQLRAYVDMQDITLNFQCIISDDKF 111
 CC DB 409 SGDRIDLYSKRAGDIQLQYG-----WKVERHI-----MDNDPVLNRPQSLHKM 453
 CC QY 112 AVLGSTEVSTQFGIGLKKFFFLSSGSADYKLOLSYENIW-----QVVLHRYPGQNAQ 164
 CC DB 454 SMMAHRVKVIPY-----STFRLNLSVTSYNADFQDGMNHLVPQSEETR 498
 CC QY 165 FLTLTOLFGAPR--IYKRLNSCYSFKEK-----PDDQWVTTTDFPFSWIG 208
 CC DB 499 AELSQCANVFLQIVSPQSNKPCMGIVQDTLCGIRKTLRDTFIELQVLNMLYVVPWDWG 558
 CC QY 209 L-----SSSLCLOFRGRVRLPNFEESFFHYAERENITLQTGTFFF----- 249
 CC DB 559 VIPTPALIKPKPLWSGKQILSVAIPNGIHLQRFDEGTLLSPKDNGLIIDGQIIFGWE 618
 CC QY 250 ----VSQKSAVNPVQPEGISIPYKILFKISSIVQHGI-----PGPALNV 292
 CC DB 619 KKTVGSSNGGLIHVVTRKPGQVCAKLFNGIQKVNFWLHNGFSTGIGTFIADGPTMRE 678
 CC QY 293 YFFRLVDPPRRR-----NVACIEHALEKLYIYKECCYD-FVRWLTSEQYDGLYKGRQ 341
 CC DB 679 IETIEAKKKVLDVTKEAQANLLTAKHGMT-----LRESFEDNVVRFLEARD-----K 728
 CC QY 342 PKKSPSTLDDGLVVRVRLVTPCKYFCPEVNVNVRNLNRSYEDID-NFLRVSYFVDEE 400
 CC DB 729 AGRLAEYNLKD-LNNVKQMYMAGSK-----GSFINIAQMSACVGGQSQSGRIAGFVDRT 783
 CC QY 401 WEKLYSTDLLPKASTGSGVTRNIYERILTLRGKFGVIGDKKFEFLAFSSQLRDNVSMVF 460
 CC DB 784 LPHFSKDDYSPE\$-----KGFV-----ENSYLR----- 806

```
QY 461 ASRFLGTAND--IRAWMGDFSQIKNVAKYARLQCSGSSRETLSVLRHEIYVDPVKVH 518
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 807 ----GLTPQEFFHAMGREGRLDTAVKTA-----ETGYIQRRVLVKALEDIMVH 851
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 519 -----GTSYVF---SDGIGISGFAH-RVASKCGLOYPSPAQIIRYGGYKGVGV 565
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 852 YDNTNRLSGNVIGFIYGEDGM-----DAAHIEKQSLDTIGGSDAAFEKRY----RVDLL 902
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 566 DPDSKMLSLKSKSYESONIKLDVLGWSKYQPCYNLRQITLL-----610
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 903 NTDHTLPSLLESSEILGD-LKLOVLLDEYKOLVDRFLRFVFDGSEANWPLPNIR 961
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 611 -----STLGVKD-----EVLQKQKEAVDQLDAI 634
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 962 RIIONAQOFTHIDTKPSDLTIKDIVLGVDLQENLLVLRLKNEIIQAQRDAVTLFCCL 1021
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 635 LHSLSKQAEALE-----LMSPGENTNLKAMLCNGYKPDAP 671
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1022 LRSRLATRRVQLQYRLTKQAFDWLNSIEAQLFSLVYHVGEMGVLA-----QSIGEP 1075
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 672 FLSMMLQTFRASKLLDLRTSRIFIPNGRTMGCGLDESRTLEYGOVVFQFTGAGHGFSD 731
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1076 ATQMTLNTFFAGVASKKVTSG--VPLRLKEITLVAKNMKT-----PSLTVYLEPGH---AA 1126
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 732 DLHPFNRSRSTNSFLIKGNVVAK-----NPLHPGDRLVLRKAVNRALHHWDCVV 784
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1127 DOEAKLIRSAIEHTTLKSVTISIASEIYDPPRSTVPIPEDEEIIQ-----LHFSLLDEE 1180
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 785 FPQKGRPH-----NECSGSDLD--GDI-----YFVCWDDQ-----MIPP 818
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1181 AEQSFQSQSPWLLRLLELDRAAMNKDLTMGOVGERIKQTFKNDLFIWSEEDNDEKLIIRC 1240
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 819 ROVQPMYPPAPSLQLOHDVTIEVEYFTNYIYVDSLGIITANAHVVFADR-----869
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1241 RWRPKSLDAETAEEDH--MLKIENTMLNITLRGNIERYVMVMKYDRKVPSPGTGEY 1298
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 870 --EPDMASDPCCKLAELFSTAVDFPKTPVPAEIPSQLRPKEYPDFMDK-----916
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1299 VKPEPWLTDGVLSEVMT-----VPGIDPTRIYNSFDIMEVLGIEAGRAALY 1349
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 917 -----PKTSYSISERVIGLFRKVKDKAPQASSIATFRDVARRSYDADMEVDGDEDY 969
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1350 KEVNVIASDGSYNYRHMLLV-----DVMFTQGGTLSTVRHFNRSNTGALMCRSFEET 1405
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 970 IDEAFDYKTEYDNKLGNDYGIKTEAILSGGIMKASKTFDRRKDAEAI 1020
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1406 VEILFEAGAS-----AELDDCRGV-SENVILGQMAPICGTGAFDVMIDEESL 1450
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 3
CTOG_HUMAN STANDARD; PRT; 2032 AA.
AC Q14008; Q14668;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CH-TOG protein (Colonic and hepatic tumor over-expressed protein).
GN KIAA0097.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain tumor;
RX MEDLINE=96128167; PubMed=8536682;
RA Charrasse S., Mazel M., Taviaux S., Berta P., Chow T., Larroque C.;
RT "Characterization of the cDNA and pattern of expression of a new gene
over-expressed in human hepatomas and colonic tumors.";
RL Eur. J. Biochem. 234:406-413(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
```

```
RX MEDLINE=95308325; PubMed=7788527;
RA Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,
RA Tabata S., Ishikawa K.-I., Kawabayashi Y., Kotani H., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. III.
The coding sequences of 40 new genes (KIAA0091-KIAA0120) deduced by
analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:37-43(1995).
CC -!- TISSUE SPECIFICITY: OVER-EXPRESSED IN HEPATOMAS AND COLONIC
TUMORS. ALSO EXPRESSED IN SKELETAL MUSCLE, BRAIN, HEART, PLACENTA,
LUNG, LIVER, KIDNEY AND PANCREAS.
CC -!- SIMILARITY: CONTAINS 9 HEAT REPEATS.
CC -!- SIMILARITY: TO C.ELEGANS F22B5.7 AND F22B5.8 AND SOME, TO YEAST
STU2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X92474; BAA63212.1; -
DR EMBL; D43948; BAA07892.2; ALT_INIT.
DR InterPro: IPR000357; HEAT_repeat.
DR PROSITE; PS50077; HEAT_REPEAT; 1.
KW Repeat.
FT REPEAT 160 197 HEAT 1.
FT REPEAT 357 394 HEAT 2.
FT REPEAT 435 472 HEAT 3.
FT REPEAT 751 788 HEAT 4.
FT REPEAT 937 974 HEAT 5.
FT REPEAT 1014 1051 HEAT 6.
FT REPEAT 1285 1322 HEAT 7.
FT REPEAT 1324 1357 HEAT 8.
FT REPEAT 1362 1399 HEAT 9.
FT CONFLICT 1564 1623 MISSING (IN REF. 1).
FT CONFLICT 1814 1814 E -> A (IN REF. 1).
FT CONFLICT 1822 1822 E -> A (IN REF. 1).
SQ SEQUENCE 2032 AA; 225507 MW; 0256603047FA45EF CRC64;

Query Match 2.1%; Score 121.5; DB 1; Length 2032;
Best local similarity 22.4%; Pred. No. 8.1;
Matches 110; Conservative 66; Mismatches 146; Indels 169; Gaps 26;

QY 276 ISSLVHGCIPGPAALNVYFPLVDPRLVPRRRNVACIEHALEKLYIKCCYDPVRMLTQYDG 335
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 565 VALIAGKGNFSKTSAQVVLGDGLVD-KIGDVKCGNNKAEMATAEACMLP--WTAEOVVS 721
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 336 YLKGROPPKSPSITLD-----DGL-----VYVRVLVTPCKV 367
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 722 MAFSQKNPKNQSETLNLWSNAIKEFGFSGNLNVKAFISNVKTALATNPVTRTAITLLGV 781
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 368 --YFCGPEVNVSNRNLNYSIEDINFLRVSVFVDEWEKL-----YST----407
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 782 MYLVGSPS-----LRFDEDEKPAL-LSQIDAEEKMGQSPAPTRGISNKHSTSGTD 833
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 408 -----DLLPKASTGSGVTRNIYERILSTLRKGFVIGDKKFEFLAIFSSSQ 451
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 834 EGEDGDEPDGSDNDVVDLLP-----RTEISDKITSELVS--KIGDKNWKI-----876
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 452 LRNSVWMFASRPGLTANDIRANWGDPSQIK-NVAKYAARLGOSFGSS-----RETLVL 505
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 877 -----RKEGL--DEVAGIINDAKFIQPNIGELPTALKGRLNDNSKILVQOITLNL 924
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 506 RH-EIEVIPDKVHGTSYVFSGIGKIS--GDFAHRYASKCGLOYPSPAQIIRYGGYKGV 552
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 925 QQLAVAMGNPKQH-----VKNLGIPIITVLGDSKNVY--RAAALATVNAAEQTGMKEWL 978
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 563 VGVDPDSSMKLSLRKSMKSYESDNKLDVLGWSKYQPCYNLRQITLLSLTLGVKDEVLEQ 622
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 979 EGEDLSEEL-----KKENPFLRQELLGW-----LAELPLTLRST--PTDLIL--1018
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```


QY 674 SMLQTFRASK-----LLDLRTSRIFIPNGRTMMGCLDESKTLEYGVQV 717
 Db 313 LMIRSCTRIADKRFMVLNLMNK-----IPD-----YCVDSKLLFEYCEV 354

RESULT 6

MOG5_CAEEL STANDARD; PRT; 1200 AA.
 AC Q09530;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable pre-mRNA splicing factor ATP-dependent RNA helicase mog-5
 DE (sex determination protein mog-5) (Masculinization of germ line
 DE protein 5).
 GN MOG-5 OR EED8.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;

[1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=20202625; PubMed=10737793;
 RA Puoti A., Kimble J.;
 RT "The hermaphrodite sperm/oocyte switch requires the Caenorhabditis
 RT elegans homologs of PRP2 and PRP22.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3276-3281(2000).
 RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Chisoe S.;

RL Submitted (JUL-1995), to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: COULD ACT LATE IN THE SPLICING OF PRE-MRNA AND MEDIATE

CC THE RELEASE OF THE SPLICED MRNA FROM SPLICEOSOMES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD

CC SUBFAMILY. DD8/PRP22 ORTHOLOG.

CC -1- SIMILARITY: CONTAINS 1 SL MOTIF DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF286899; AAC01332.1; -

DR EMBL; U23484; AAC46765.1; -

DR HSSP; P05055; ISRO.

DR WormPep; EED8.5; CE01889.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002464; DEAD_BOX.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR003029; Sl.

DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00575; Sl; 1.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELIC; 1.

DR SMART; SM00316; Sl; 1.

DR PROSITE; PS00690; DEAD_ATP_HELICASE; 1.

DR PROSITE; PS0126; Sl; 1.

KW mRNA processing; mRNA splicing; Helicase; ATP-binding;

KW Nuclear protein; Developmental protein.

FT DOMAIN 176 185 POLY-ARG.

FT DOMAIN 234 304 SL MOTIF.

FT NP_BIND 563 570 ATP (POTENTIAL).

FT SITE 660 663 DEAD_BOX.

SQ SEQUENCE 1200 AA; 135759 MW; 023C46E494D7F4F CRC64;

Query Match 2.08; Score 116.5; DB 1; Length 1200;
 Best Local Similarity 18.88; Pred. NO. 7.9;
 Matches 166; Conservative 137; Mismatches 331; Indels 247; Gaps 44;

QY 328 WLTEQYDGLKGRQPPKPSIITLDDGLVYVRRVLVTP-----CKYVFG-----PEVN 375
 Db 340 WNPESAGVGQG-----PSTSIGGKARVR--ISTPERWELRQMOGAGVLTATDMPDFD 391
 QY 376 VSNRVLNYSIEDID-NFLRVSVFVDEWEKLY-----STDLLP-----KASTGSGVRTNIYE 425
 Db 392 EEMGLRNYYDDESDGEDIIELVDEPDPLRGYKGGAIEPVKVVKNPNDPGLAALMQ 451
 QY 426 RILSTLRKGFVIGDKKFEFL-----AFSS--SOLRDNSSVMFASRPGLITANDIRAMGDFSQ 480
 Db 452 GALSKEKRTKTOAQERDMDTQKGFSSNARILD-----PMSGNQSAWSADESK 501
 QY 481 IKN-----VAKYAARLQSGSSREYLSVL---RHELEIVPDVKVHGT 520
 Db 502 DRNNKMKEMPEWLKHVYTAGGKATYGRRTNLSMVEQRESLPIFALKKNLMEAMIDNQI--L 559
 QY 521 SVFSDGIGKISGDFAHRVASKGQYTPSAFQIRYGYGVGVG-VDPSSMKLSLRKSM 579
 Db 560 VVVGEGSGKTT-----QMTQYALEAGL-GRRGKIGCTQPRVAAMSAVKRV 605
 QY 580 SKYESDNIKDLVMSKYQPCVNLNQLITLSTLGVKDEVLEQKQKEAVDQDLALHDSL 639
 Db 606 ABEYGCGLGTGVYTIREFDCTSDTIIKYMT-----DGML---LRECLIDPDLGYSLI 657
 QY 640 KAAQALELMSPGENTNLIKAMLCNGYKPDAPFLSMMLQTFRASKLLDLRTSRIFPNG 699
 Db 658 MIDEAHERTI---HTDVLGLKAAARKPELKIITSLTDSVKEFSEFLEAPITFIPG 714
 QY 700 RTMMGCLDESRTLEYGQVQVFTGAGHGFSDDLHPFNNSRSTNFIKNGVYVAKNCP 759
 Db 715 RTF-----PVEILYT-----REPESDYLEAAHITVMQ--- 741
 QY 760 LH-----PGDIRVL---KAVNVRALHVMVDCVVFQPKGRPHNECSGSDLDGDIYFVCW 811
 Db 742 IHLTEPPGDLVFLTGQEEID-----TSCEVLIERMK-----SMGPDVPELI----- 783
 QY 812 DQDMIP-----PRQVQPMPEYPAPSIQLDHDVTIEEVEEFTNYIVNDSGLIIANHVVF 866
 Db 784 ---ILPVYALPSEMOTRIFDPAPAGK--RKVVIA-----TN-IAETSLTIDGIFYV- 830
 QY 867 ADREPDMAISDCKLAEILFSTAVDPKGTGPAEIPSQLRPKEYPDFMDPKDKTSVISER 926
 Db 831 -----DQGVKQKIYN-----PRSGMDSLVTVPISQAAAKQSRGRGTG----- 870
 QY 927 VIGKLFPRYKDKAPQASSIATFTRDVARSYDA---DMEVDGFDYIDFAFYKTEYDN- 982
 Db 871 -PGKCYRLYTERAFRDEMLPTVPVEIQRNTLASTLIQLKAMGINLLIDPDMADPLDSM 929
 QY 983 -----KLGNDYVYGIKTEAEILSGGIMKASKTFRDKD---AEA 1019
 Db 930 ITALNTLHTLSALDGDGLTKLGRNNAEPLEP-----SUSKLLIMSVDLGCSEE 979
 QY 1020 ISVAVRALRKEARAFKRRNDIDMLPKASAWYHVTY--HPTYWGCYNOGLKRAHFISFP 1077
 Db 980 VLTIVAML--NVQNIYFRPKQDHADQKAKFQHPGEGDHLTLAYVNSW--KNHHSQP 1035
 QY 1078 WCYVDQLIQIKDKKARNRPVNLSSLAQL-----SHRLVL 1113
 Db 1036 WC-FENFIQVRSMK-----RAQDIRKOLLGIMDRHKLLM 1068

RESULT 7

ORPL_HUMAN STANDARD; PRT; 950 AA.
 ID ORPL_HUMAN Q9BNW8; Q9B2F5;
 AC Q9BNW8; Q9BNW8; Q9B2F5;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Oxysterol binding protein-related protein 1 (OSBP-related protein 1)
 DE (ORP-1).
 GN OSBPL1 OR ORP1 OR OSBP8.
 OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RX MEDLINE=21601154; PubMed=11735225;
RA Jaworski C.J., Moreira E., Li A., Lee R., Rodriguez I.R.;
RT "A family of 12 human genes containing oxysterol-binding domains.";
RL Genomics 78:185-196(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=21264638; PubMed=11279184;
RA Xu Y., Liu Y., Ridgway N.D., McMaster C.R.;
RT "Novel members of the human oxysterol-binding protein family bind
RL phospholipids and regulate vesicle transport.";
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM C).
RX MEDLINE=21376257; PubMed=11483621;
RA Lahtonen M., Laitinen S., Chinetti G., Johansson M., Ehnholm C.,
RA Staelens B., Ikonen E., Olkkonen V.M.;
RT "The OSBP-related protein family in humans.";
RL J. Lipid Res. 42:1203-1213(2001).
RN [4]
RP SEQUENCE OF 514-790 FROM N.A. (ISOFORM A).
RC TISSUE=Embryo;
RA Ito T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: Binds phospholipids; exhibits strong binding to
CC phosphatidic acid and weak binding to phosphatidylinositol 3-
CC phosphate (by similarity).
CC 2- ALTERNATIVE PRODUCTS: 3 isoforms; A/OSBP1A/OSBP8S,
CC B/OSBP1B/OSBP8L (shown here), and C; are produced by alternative
CC splicing.
CC 3- SIMILARITY: BELONGS TO THE OSBP FAMILY.
CC 4- SIMILARITY: CONTAINS 3 ANK REPEATS.
CC 5- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF392449; AAL40662.1; -
DR EMBL; AF392450; AAL40663.1; -
DR EMBL; AF274714; AAK15154.1; -
DR EMBL; AF323726; AAG53407.2; -
DR EMBL; AK001079; BAA91496.1; -
DR Genbank; HGNC:16398; OSBP1A.
DR MIM; 606730; -
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000648; Oxysterol_BP.
DR InterPro; IPR001849; PH.
DR Pfam; PF00023; ank; 4.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF01237; Oxysterol_BP; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS01013; OSBP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Lipid transport; Transport; ANK repeat; Repeat; Alternative splicing.
FT REPEAT 47 76
FT REPEAT 80 109
FT ANK 1.
FT ANK 2.

FT REPEAT 175 204 ANK 3.
FT DOMAIN 235 334 PH.
FT VARSPLIC 1 513 MISSING (IN ISOFORM A).
FT VARSPLIC 625 648 FNPLIGTYELVRDGLFRLISEQ -> L (IN ISOFORM C).
FT CONFLICT 735 735 T -> A (IN REF. 4).
FT CONFLICT 810 810 S -> P (IN REF. 1; AAL40663).
FT CONFLICT 841 841 P -> S (IN REF. 1; AAL40662).
FT CONFLICT 844 844 A -> G (IN REF. 1; AAL40662).
SQ SEQUENCE 950 AA; 108469 MW; CDDA26CA27B65F63 CRC64;
Query Match 2.0%; Score 116; DB 1; Length 950;
Best Local Similarity 18.9%; Pred. No. 6;
Matches 116; Conservative 101; Mismatches 234; Indels 164; Gaps 31;
QY 592 LGWSK-YOPCYL-NRQLTLLSTLVGVKDEVLEQKQKAEVADQDALIHDLSL---KAQEALE 646
DB 47 LGWTPHLACVYFGRVQVVDLLKAGAEVNLN-----DMGDTPLHRAATGRKELVNL 99
QY 647 LMSPGENTNI-----LKAMLNGCYKPDAPFLSMMLQTFRASKLLDURT 690
DB 100 LLEYNADTIVNGSQTAKEVTHAEIRSMLEAVERTQOKLEELLAAAREGKTTEL-- 157
QY 691 RSRIFIPN-----GRVMGCLDESRTLEYQGVVQVFTGAG---HGFESDDLHPFN 737
DB 158 TALLARPNDPVCSDQLGNTFLHC---AAYRAHKQCALKLLRSAGADPNLKNKNDQKPLD 214
QY 738 NSRSTNSNFIKGNVVAKNPCLHPGDI-----RVLKAVNVYRALHHMV----- 780
DB 215 LAQGAEMKHLVGNVVIKALRYEGPLWKSRSFEGWLFVWV---LEHGVLSWYRKOP 270
QY 781 DGV--VFPOKGRPHPNCEGSDLDGDYFV-CWDQDMIPPRQVQPMYPPAPSIQLOHD 837
DB 271 DAVHNIYRGCKHLTQAVCTVKSTDSCLFFIKCFD-DIHHGFRV-----PKNSLQOSRE 323
QY 838 VTIEVEE---YFTNVVNDSLGIITANAHVFPADREPDMAMSDPCCKLAELFSIADVPFK 894
DB 324 DWLEAIEESASVTHYCSQDL-----TDEEEDTVS-----AALKK 361
QY 895 TGVPAEIPSQLRPKEYPDFMDKPDXTYSISERVIGKLFKVK--DKAPQASSIA----- 946
DB 362 SLEKAQSCQOQLRDRETSNFL-KMIKEDCMAKEMLPFLQKVEVWSEASRETCAVLTDCIN 420
QY 947 TTRDVARSYDADMEVDGFEDYIDEAFD-YKTEYDNKLGMLDYVYIKTEAIIISG--- 1002
DB 421 LFTKQBGVNFLEQEQER-KNILSEALETLATEHH-----ELEQSLVKGSP 467
QY 1003 -GIMKASKTFDRKDAEA-----ISVAVRALRKARAWFRNRNDIDDMLPKASAWHV 1054
DB 468 ASTLSDEFYDALSDSESESRSLRLEAVTARSTFEFEHGLSGRKHMSSEKDCGGG--- 523
QY 1055 TYHTPTWGCYNQGLKRAHFSPP-----W-----CVYDOLIOIKDKARNRVL 1098
DB 524 -----DALSGIKK-HRTSLPSPNFSRNDPSIILKRCIGMELSKITMPVINEPLS 575
QY 1099 NLSSLAQLSHRLVL 1113
DB 576 FLQRLTEYMEHTYLI 590
RESULT 8
PYRL_YEAST
ID PYRL_YEAST STANDARD; PRT; 2214 AA.
AC P07259;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE URAL protein [includes: Glutamine-dependent carbamoyl-phosphate
DE synthase (EC 6.3.5.5); Aspartate carbamoyltransferase (EC 2.1.3.2)].
GN URA2 OR YJL130C OR J0686.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 28583 / FL100;
 RX MEDLINE=89378778; PubMed=2570735;
 RA Souciet J.-L., Nagy M., le Gouar M., Lacroite F., Potier S.;
 RT "Organization of the yeast URA2 gene: identification of a defective
 RT dihydroorotase-like domain in the multifunctional carbamoylphosphate
 RT synthetase-aspartate transcarbamylase complex.";
 RL Gene 79:59-70(1989).
 RN [2]
 RP SEQUENCE OF 1-510 FROM N.A.
 RX MEDLINE=87286375; PubMed=3039294;
 RA Souciet J.-L., Potier S., Hubert J.-C., Lacroite F.;
 RT "Nucleotide sequence of the pyrimidine specific carbamoyl phosphate
 RT synthetase, a part of the yeast multifunctional protein encoded by
 RT the URA2 gene.";
 RL Mol. Gen. Genet. 207:314-319(1987).
 RN [3]
 RP SEQUENCE OF 1268-2214 FROM N.A.
 RX MEDLINE=89255278; PubMed=2498313;
 RA Nagy M., le Gouar M., Potier S., Souciet J.-L., Herve G.;
 RT "The primary structure of the aspartate transcarbamylase region of
 RT the URA2 gene product in Saccharomyces cerevisiae. Features involved
 RT in activity and nuclear localization.";
 RL J. Biol. Chem. 264:8366-8374(1989).
 RN [4]
 RP SEQUENCE OF 1-276 FROM N.A.
 RX STRAIN=S288c / FY1679;
 RC MEDLINE=97103775; PubMed=8948101;
 RA Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
 RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
 RT reveals 19 open reading frames including URA2 [5' end], TRK1, PBS2,
 RT SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, COT7, Gzf3, two tRNA genes,
 RT three remnant delta elements and a Ty4 transposon.";
 RL Yeast 12:1471-1474(1996).
 RN [5]
 RP SEQUENCE OF 175-2214 FROM N.A.
 RX STRAIN=S288c / FY1679;
 RC MEDLINE=96408771; PubMed=8913765;
 RA Katsoulou C., Tzermia M., Tavernarakis N., Alexandraki D.;
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast
 RT chromosome X reveals 14 known genes and 13 new open reading frames
 RT including homologues of genes clustered on the right arm of
 RT chromosome XI.";
 RL Yeast 12:787-797(1996).
 CC -1- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING THREE
 CC ENZYMATIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSASE,
 CC AND ATCASE).
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + glutamate + carbamoyl phosphate.
 CC -1- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
 CC + N-carbamoyl-L-aspartate.
 CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
 CC -1- PATHWAY: Pyrimidine biosynthesis; second step.
 CC -1- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSASE
 CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
 CC GLUTAMINE-DEPENDENT CPSASE (GD-CPSASE) (EC 6.3.5.5).
 CC MISCELLANEOUS: IN EUKARYOTES EC 6.3.5.5 IS SYNTHESIZED BY TWO
 CC PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.
 CC -1- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPSASES.
 CC -1- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE DHOASE FAMILY.
 CC DEFECTIVE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M27174; AAA68280.1; ALT_INIT.
 DR EMBL; X05553; CAA29068.1; -
 DR EMBL; J04711; AAA35198.1; -
 DR EMBL; Z49405; CAA89425.1; -
 DR EMBL; X87371; CAA60825.1; -
 DR EMBL; D28139; BAA05680.1; -
 DR PIR; S05767; Q2BYU2.
 DR HSP; P00968; I49X.
 DR SGD; S0003666; URA2.
 DR InterPro; IPR002029; Asp/Oro_Cotranf.
 DR InterPro; IPR002082; Asp_carbmltransf.
 DR InterPro; IPR001317; CPS_GATase.
 DR InterPro; IPR000901; CPSase.
 DR InterPro; IPR002474; CPSase_sm_chain.
 DR InterPro; IPR002195; Dihydroorotase.
 DR InterPro; IPR000991; GATase_1.
 DR InterPro; IPR004362; MGS_like.
 DR Pfam; PF00117; GATase; 1.
 DR Pfam; PF00185; OTCace; 1.
 DR Pfam; PF00289; CPSase_L_chain; 2.
 DR Pfam; PF00744; Dihydroorotase; 1.
 DR Pfam; PF00988; CPSase_sm_chain; 1.
 DR Pfam; PF02142; MGS; 1.
 DR Pfam; PF02729; OTCace_N; 1.
 DR Pfam; PF02786; CPSase_L_D2; 3.
 DR Pfam; PF02787; CPSase_L_D3; 1.
 DR PRINTS; PRO0100; AOTCASE.
 DR PRINTS; PRO0098; CPSASE.
 DR PRINTS; PRO0099; CPSGATASE.
 DR PRINTS; PRO0096; GATASE.
 DR TIGRFAMS; TIGR00670; asp_carb_tr; 1.
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
 DR PROSITE; PS00442; GATASE_TYPE_1; 1.
 DR PROSITE; PS00866; CPSASE_1; 2.
 DR PROSITE; PS00867; CPSASE_2; 2.
 KW Pyrimidine biosynthesis; Ligase; Transferase; Multifunctional enzyme.
 FT DOMAIN 1 400
 FT DOMAIN 401 440
 FT DOMAIN 441 1482
 FT DOMAIN 1483 1492
 FT DOMAIN 1493 1821
 FT DOMAIN 1822 1909
 FT DOMAIN 1910 2214
 FT ACT_SITE 302 302
 FT ACT_SITE 386 386
 FT ACT_SITE 388 388
 FT CONFLICT 86 86
 FT CONFLICT 123 123
 FT CONFLICT 250 257
 FT CONFLICT 270 270
 FT CONFLICT 313 314
 FT CONFLICT 372 373
 FT CONFLICT 394 402
 FT CONFLICT 431 433
 FT CONFLICT 482 482
 FT CONFLICT 485 485
 FT CONFLICT 492 492
 FT CONFLICT 501 510
 FT CONFLICT 502 852
 FT CONFLICT 852 852
 FT CONFLICT 1411 1412
 FT CONFLICT 1482 1482
 FT CONFLICT 1582 1588
 FT CONFLICT 1592 1592
 FT CONFLICT 1595 1595
 FT CONFLICT 1937 1937
 FT CONFLICT 1937 1937
 FT CONFLICT 1997 1997
 FT CONFLICT 2039 2039
 FT CONFLICT 2158 2165
 FT CONFLICT 2214 AA; 245124 MW; 4CA58304DRECAD21 CRC64;
 SQ
 Query Match 2.0%; Score 114.5; DB 1; Length 2214;
 Best Local Similarity 18.8%; Pred. NO. 28;
 Matches 153; Conservative 119; Mismatches 267; Indels 273; Gaps 43;

```
QY 384 YSEDIDNF-----LRVSFVDEEHEKLYSTDL-----LPKASTGSGVRTNIYERILST 430
DB 347 FAVDVTITTSQWKLFLVNANDDSNEGIYHSELPSYFSVQHPESIPG-----PRTEF 398
QY 431 LRKGVIGDKKEF-----LAFSSQLRNSVMFASRFGLTANDIRA-WMGDFSQIKN 483
DB 399 LFDVFIQAVKEFKYTVLKPFAFPGGLEDN---VRAHPRIEAKKVLVLGSGGLS---- 450
QY 484 VAKYAARLQSGFSGSSRETLVLRIE---TEVIPDKVHGTYSYFSDGI--GKISGDFAH 537
DB 451 ----IGGAGEFDYGSQAIRKALKEEGYITILINPNIATISKGLADKVFPVPTAEFVR 506
QY 538 RVASKGGLQYTPSAFQIRYGYKGV-VGVDPDSMKLSLRKMSKSYSDNIKLDVLGWSK 596
DB 507 KVI---LHERPDALYVFGQTALSVGI---AMK-----DEFALGVKV----- 544
QY 597 YQPCYLNRLITLST-----LVKDEVLQOK-K-EAVDQDAILHDSLKAQFALEMS 649
DB 545 ----LGPDTITITTEDRELFNSAIDINEKCAKSAQANSVO-----EALAAVK 589
QY 650 PGENTNLIKAM-----LNCGYKYPDAEPFLSMQLQTFRASKLLDLRTSRIFIPNRTMMG 704
DB 590 EIGFPVIVRAAVALGGLSGFANKEKELVDLCNVAFSSP-----EALAAVK 589
QY 705 CLDESRTLEYQVQVQFTGAGHGFSDDL--HPENN--SRSTNSNFILKG-----NVVVA 755
DB 630 -----QVLVEKSMKWKVEYEVVDAFDCNITVCNMENFDPLGHTGDSIWA 678
QY 756 KNPCLHPGDIRVLKAVNRALHMM-----VDCVFPQKGRPHNECS 798
DB 679 PSQTLSDEDYNMLRTAVNVIHLGVYCECNIQYALNPVSKDYCII--EYNARLSRSAL 736
QY 799 GSDLDGDIYFCWODMTTPRQVQPMYPPAPSIQLDHDVTIEVEEYFTNIYVNDSLGI 858
DB 737 ASKATG-----YPLAYTAA---KLGLNIPLNEVKSNTK----- 767
QY 859 IANAHVFAADREPDAMSDPCKKLAELFSIAVDPPK-----TGVAEIPSQLPKEVPD 912
DB 768 -----STACPEPSL---DYC-----VVKPRWDLKFTTRVSELSSM--KSVGE 808
QY 913 FMDKPKDTSYISERVIGLFRKVKDKAPQASIAFTTRDVARRSYDADMEVDGDEY--- 969
DB 809 VMS-----IGRTFEEAIQKAIRSTEYANLGFNETDLIDIDYELNPNPTDMRVF 856
QY 970 -IDEAFDYK-----TEYD-----NKLGNDY-----YGIKTEAETLSGGIMKAS 1008
DB 857 ATANAFKGYSDVKWEMTRIDKWLKHLDLVQFAEKISFGUKELPSEL---VLKQA 913
QY 1009 KT--FDRRKDA---BAISVAVRALKE--ARAWFKRRNDIDDMFLPKASAWYHVTY---H 1057
DB 914 KQLGPDQRQIARFLDSNEVAIRLRKEYGITPFVKQIDTVAAEFPAYTNILYMTYNADSH 973
QY 1058 PTYW-----GCYNOGLKRAHFISFPWC 1079
DB 974 DLSFDDHGYMVLGSGYIRIGSS---VEFDWC 1001
```

RESULT 9

```
SYA_THETN STANDARD; PRG: 878 AA.
ID Q8RAH4;
AC DT 15-JUN-2002 (Rel. 41, Created)
OC DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine-tRNA ligase) (Alars).
OS ALAS OR TIE1248.
GN Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xie Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.,
RA A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
CC -L- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -L- SUBCELLULAR LOCATION: Cytoplasmic.
CC -L- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE013087; AAC24472.1;
CC PROSITE: P50860; AA_TRNA_LIGASE_IL_ALA; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC
CC SEQUENCE 878 AA; 98363 MW; F2CE98288DC2A244 CRC64;
```

Query Match 1.9%; Score 113.5; DB 1; Length 878;
Best Local Similarity 19.4%; Fied. NO. 7.9; 308; Indels 327; Gaps 46;
Matches 180; Conservative 113; Mismatches

```
QY 329 LTRQYDGYLKGROPKPSITL-----DGLVYVRRYLVTPCKYVFCGPEVNSRV--- 380
DB 9 IREKFLSFPESKGLRLPSFLPKNDKSLILNSGM-APLKPYPFTGKETPPRRVTTCC 67
QY 381 -----LRNYS-----EDIDNLRVSVFDEWEK 403
DB 68 RCITPDIERVGTARHATFEMLGNFSGDYFKKKAIPMAWEFAEVLKLP-VDRLWVT 126
QY 404 LYSTD-----LLPKASTGSGVRTNIYERILSTLRK-----FVIGDKK----- 441
DB 127 IYEDDEAFIWNKIVGLPPIRIVRMKEDNFEW--IGTGPCGPGCEIYFDRGEKCGK 184
QY 442 -----FEFLAFSSQLRNSVMFASRP-----GLTANDIRAWNG----- 476
DB 185 PDGIGIGDCCDRVFEWNLVFTQFNKDEQNYHRLPNPNTDTGMGLERIAIMQGVDTIF 244
QY 477 DFGSIKNVAKVAAIRGO-SFGSSRETLVLRHEIETIPDVKVHGTYSYFSDGI----- 528
DB 245 DVDVIRGIRDFISDLAEVEYKDAK---DVSIRVITD-HIRGITFMSDGLILPSNEGR 299
QY 529 GKISGDFAHRVASKGGLQYTPSAFQIR-----YGG-YKGVVGVDPDSSMKLSRKSM 579
DB 300 GYVLRLLRRAARHGKLLGLNDAFLYKVDVSVVENTGGAYPEII-----ER 345
QY 580 SKYESDNIKLDVLGWSKYQPCYLNRLITLLSTLGVKDEVLEQKQEAVDQLDAIILD-- 637
DB 346 KDYIKKIILE-----EERFETVDQGLAILQDI 375
QY 638 -SLKAQ-----EALMLSP-GENTNLIKAML-----NCGYKPDAPFLSMMLQT 679
DB 376 NELKAQGTIVLEGAKAFKLYDTYGFPLDLAKELIQESGITIDEEGYKEELEK-----QR 429
QY 680 FRA-----SKLLDLRTSRIFI-----PNGRMTMGCLESRTLE----- 713
DB 430 VRASRKRKEDNSLMEQDIYSTGLDISTK---FVGYEYSEAKVLAIVKNEVVEQAEAG 486
QY 714 -----YQQVFVQFTGAGHGFSDDLHPFNNSRSTNSNFILKGNV----- 752
DB 487 DDVSIILLDVTFFYAESGGQVSDSVIEEDTLIKVNDCKKVKGNKFIHIGITVERGLISVGD 546
QY 753 -VWAKNCPCLHPGDIRVLKAVNRALH---HMVDCVFPQKGRPHNECSGLDGDYIF 808
DB 547 KWAK-----IDVSKRKGAAARNHTATHLHLKALKEVLG--DHVWQ-SGS-----L 588
```


Query Match	1.9%;	Score 113;	DB 1;	Length 787;
Best Local Similarity	19.1%;	Pred. No. 7.2;		
Matches 139;	Conservative 113;	Mismatches 270;	Indels 206;	Gaps
Qy 372	PEVNVSNRVURN---	YSEDIDNFLRVSFVDEEKEKLYSTDLLPKASTGSGVRTNI	VERIL 428	
	: : : : :	: : : : :	: : : : :	
Db 198	PIILNGLFKISNTIFGRTSDEKIKYFESMNEWEPF-----	VEGDDEPKVH 244		
Qy 429	STLRKGVVIGKKPEFLAFSSQJLRNS-----VW---	MFASRPGLTANDIRAWMGDFSQ 480		
	: : : : :	: : : : :	: : : : :	
Db 245	PALAKAM--DEAVEKIKAIQKNARENDATLEPWPWMI	VFRAPKGWTGP--KSWDGD---	296	
Qy 481	IKNVAKYAARLGOSSGSSRETLFV---LRHELEVPDVKVHGTSV	VFSDGIGKISGDEFA 536		
	: : : : :	: : : : :	: : : : :	
Db 297	-----KIEGSFRAHQIPIVDQTDMEHADALVDWLESYQPKLFNED	GLSKDDIK 346		
Qy 537	HRVASKGGLGYTSPAQIRVGGYKGVVGDPPDSMSKLSLRK---	SMKYESDNIKLDVLG 593		
	: : : : :	: : : : :	: : : : :	
Db 347	EIIPTG--DARMAANPITG-----GVDPKALNLFNFRDRAVDTSKGA-	NRVODMTV 396		

[illegible]


```
Db 130 LVSTFNVLSILESKNQ-----LLEIGSDAMHVLLSGK-SIFEALPN 174
QY 194 DQWVTTDFP-----PSWIGLSSLCIOFRGV 221
Db 175 DNYLQISGLPKNNVFTVSKKRRITERTITONKARKEVWNSISIRFIFRYS 234
QY 222 RLPNFEESFFHYAERENITLTGTFFVSQKSLVNPVQPEGISIPYKILFISSLV- 280
Db 235 YKFKQDLVFNLSICDRNTVHMWLOWFPQFGLNFAQ-----VKQHKVPIYS 286
QY 281 QHGCIPTGALNFFRLVDRRRNVACIEHALEKLYIYKECCYDPVRLWLTBOYDYGKGR 340
Db 287 QSTVVPKRLKYV--PLIEQAKRLERI--SLSKYV-----NHYCPYIDTH 328
QY 341 QPPKSPSTLDDGLV--VYRRVLVTPCKVFCGPRVNSN-RVLNYSIEDTDLRLVSEV 397
Db 329 DDEKILSYSLAPNQVAFELRSILV--RVF---FKLWGNQRIEILKDLTFLKLSRY 382
QY 398 DE-----EWEKLYSTDLLPKASTGSGVRTNIYERILSTLRKFGVIGDKKF 442
Db 383 ESFSLHYLMSNIKISEIWLVLGRSNNAKMLSDPEKRRQIFAEBFYWLYNSFIPILOS 442
QY 443 EFLAFSSSOLDNSV-----WMPASRPLGTANDIRAW----- 474
Db 443 FFIYTESDLNRRTVYFRKDIWKLCLRPFTTSMKMEAFKINENNVRMDTKTLPPAVI 502
QY 475 -----MGDFSQIKNAK-YAARLGOS---FGSSRETL---SVLRH---EIEVIP-DVK 516
Db 503 RLLPKKNTFRITLNRKRLIKWGSNKKMLYSTNQTLPVASILKHLNESSGIPFNLE 562
QY 517 VHGTSYVSDGKIGKISDFAHVASKGLQVTPSAFQIRYGGYGVGVDPDSSMKLSLR 576
Db 563 VYKMLLTFFKKDLK-----HRM-----FGKKYFVRIDIKSY----- 595
QY 577 KMSKYESDNIKLD---VLGWSKYQPCYLNRLQTLITLTLGVXDEVLEQKQKAVDQLD 632
Db 596 -----DRIKQDLMFRLVKKLKDPEVIRKVIATHT---SDRATKNVSEAFSYFD 644
QY 633 AILHDSLKAQEALEMLSPGENTNLIKAMNGYKPDAPFFLSMMLQTFRASKLIDL--- 689
Db 645 MY-----PFEKVQLLSMKTSITLFDVFDVYWKSSSEIF--KMLKEHLSGHIVKIGNSQ 697
QY 690 TRSRIFIPNGRT-----MMGCLDE--SRTLEYGOVFEQ 720
Db 698 YLQKVGIQFGSILSFLCHFTMEDLIDEVLSFTKKKGSVLLR 739

RESULT 13
ID DPOOL_HELPY
AC P56105;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA OR HP1470.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26595 / ATCC 700392;
RX MEDLINE=97394457; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathley L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
```

```
Venter J.C.:
"the complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RL Nature 388:539-547(1997).
CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N)
CC -!- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000647; AAD08510.1; ALT_INIT.
CC HSP: P19821; 1TAU.
CC TIGR: HP1470.
CC
CC InterPro: IPR002562; 3_5_exonuclease.
CC InterPro: IPR002421; 5_3_exonuclease.
CC InterPro: IPR001098; DNA_pol.
CC InterPro: IPR002298; DNA_pol.
CC InterPro: IPR000513; Exo_N.1.
CC InterPro: IPR003584; HHH.2.
CC Pfam: PF00476; DNA_pol_A.1.
CC Pfam: PF01367; 5_3_exonuclease; 1.
CC PRINTS: PR00868; DNAPOLI.
CC SMART: SM00474; 3SEXOC; 1.
CC SMART: SM00475; 5SEXOC; 1.
CC SMART: SM00279; HHH2; 1.
CC SMART: SM00482; POLAC; 1.
CC TIGRfams: TIGR00593; pole; 1.
CC PROSITE: PS00447; DNA_POLYMERASE_A; 1.
CC
CC Transferrase; DNA-directed DNA polymerase; DNA replication; DNA repair;
CC Hydrolyase; Exonuclease; DNA-binding; Complete proteome.
CC
CC SEQUENCE 891 AA; 101975 MW; 3D6E7D8FF613EB88 CRC64;

Query Match 1.9%; Score 111.5; DB 1; Length 891;
Best Local Similarity 18.5%; Pred. No. 11;
Matches 138; Conservative 114; Mismatches 231; Indels 261; Gaps 38;

QY 402 EKLY-STDLLPKASTGSGVRTNIYERI-----LSTLRKFGVIGDKKFEFL--AFSS-- 449
Db 213 EKIVENDLAKNLLSPKNYRALTHDKASAFLSKELATLARGCI---KEFDLSCAFPSEN 269
QY 450 -----SOLRD-----NSVWMEASRPLGTANDIRAMWGDFOIKNV 484
Db 270 PLIKIKDELKEYGFTSTLRDLNENSPPTLIIDNAPLDNTPALDNTPKKSCM-----IV 322
QY 485 AKYAAALGQSGSGSRRTLSVLRIHEIPIPVKVGTSYFVSDDGIGTSGDF----AHRVAS 541
Db 323 LESAAPLASFLEKLEKTARVAFARLVLDKKEKVALALAFYED----QGYPLEEALES 377
QY 542 KGLQYTPSAFQIRYGGYGVGVGVDPSDKLSRMSKSYESDNTKLDVLGSKYQPCY 601
Db 378 PFSLEFLQNAF-FKMLQHAQIIG--HDLKPLLSFLSKAKYQVPLENIRIQ----- 423
QY 602 LNROLITLLST---LGVKDFVLEQKQKQKAVDQLDAILHDSLK---AQEALEMLSPGEN- 653
Db 424 -DTQILAFLNPKVGF-DEVLEKYLE-----ELPHEKIKDFKPKAEKLELLSVELNA 476
QY 654 -----TNILKAMLCN---GYKPDAPFPLSMMLQTFR---- 681
Db 477 LKRLCEYFEKGGLEENLLSLARBIETPFMKVLGMGMEFQGFKIDA-PYFKLEQEFKNEH 535
QY 682 -----ASKLLDLRHSRIFIPNG-----RTMMGCLDE---SRTLE 713
```

Db 536 VLQRILLELIGVDFNLNSPKOLSEVLYDKLGLPKNKSHTDEKSLKILDKHPSIALILE 595
 QY 714 YGOV-----FVOFTGAGHGFSDDLHPNPN--SRSTNSNFI 747
 Db 596 YRELKLFNTYTTLLRLKDKKDDKHHTFIQ-TGTATGRLLSSHPNLQNPVRSPPKGLLI 654
 QY 748 LKGNVWVAKNCPCLHPGDIRVLKANVRALHVMVDCVVFPQKGRPHPNECSSGSLDGDYI 807
 Db 655 RKGLIASKEVCLLQVD---YSQELRLAH-----FSQ----- 685
 QY 808 FVCWQDMIPPRQVOPMEYPPAPSIQDHDVTHIEVEYFTNYIVNDSLGIIANAHVFFA 867
 Db 686 ----DKDLM-----EAFKGRDHLFTSKALFGEVLAKEKRSIAKSINP--- 725
 QY 868 DREPDMAWSDPCKKLAEELFSTAVDFPKTGVPAEIPSLRPEKYPDFDKDKTSYISERV 927
 Db 726 ----GLVYGMGSKSLSTLNLSNEAKSYEAYF-----KRPSTIKD-----YLM--- 766
 QY 928 IGLKFRVKDKAPQASSIAFTTRDVARSYDADMEVDGFEDYIDEADPYKTEYDNKLGNL 987
 Db 767 -----RMKEEILKTSK--AFTLLGRVRFVDF---FTGANDYVKG--NYLREGVNAI--- 809
 QY 988 MDYVGIKTEAELISGGIMKASKTF 1011
 Db 810 --FOG--SASDLLKGLMKVRSERF 829

RESULT 14
 ID LAMA4_HUMAN STANDARD; PRT; 1816 AA.
 AC Q16363; Q15335; Q14735; Q9UE18; Q9JUN9;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-4 chain precursor.
 GN LAMA4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP TISSUE=Fetal lung;
 RC MEDLINE=95300971; PubMed=7781776;
 RX Iivanainen A., Sainio K., Sariola H., Tryggvason K.;
 RA "Primary structure and expression of a novel human laminin alpha 4
 chain.";
 RT FEBS Lett. 365:183-188(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=97434279; PubMed=9310354;
 RA Richards A.J., Luccarini C., Pope F.M.;
 RT "The structural organisation of LAMA4, the gene encoding laminin
 alpha4.";
 RL Eur. J. Biochem. 248:15-23(1997).
 RN [3]
 RP SEQUENCE OF 236-1816 FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=95048381; PubMed=7959779;
 RA Richards A.J., Al-Imara L., Carter N.P., Lloyd J.C., Leversha M.A.,
 RA Pope F.M.;
 RT "Localization of the gene (LAMA4) to chromosome 6q21 and isolation of
 a partial cDNA encoding a variant laminin A chain.";
 RL Genomics 22:237-239(1994).
 RN [4]
 RP SEQUENCE OF 66-1816 FROM N.A.
 RA Tubby B.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.

CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 CC different polypeptide chains (alpha, beta, gamma), which are bound
 CC to each other by disulfide bonds into a cross-shaped molecule
 CC comprising one long and three short arms with globules at each
 CC end.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 CC MEMBRANES (MAJOR COMPONENT).
 CC -!- TISSUE SPECIFICITY: IN ADULT, STRONG EXPRESSION IN HEART, LUNG,
 CC OVARY SMALL AND LARGE INTESTINES, PLACENTA, LIVER; WEAK OR NO
 CC EXPRESSION IN SKELETAL MUSCLE, KIDNEY, PANCREAS, TESTIS, PROSTATE,
 CC BRAIN. HIGH EXPRESSION IN FETAL LUNG AND KIDNEY. EXPRESSION IN
 CC FETAL AND NEWBORN TISSUES IS OBSERVED IN CERTAIN MESENCHYMAL CELLS
 CC IN TISSUES SUCH AS SMOOTH MUSCLE AND DERMIS.
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAIN G IS GLOBULAR.
 CC -!- SIMILARITY: CONTAINS 3-5 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -!- CAUTION: GENE LAMA4 WAS FORMERLY CALLED LAMA3.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; S78569; AAB34635.1; -
 CC EMBL; X91171; CAB62596.1; -
 CC EMBL; Y14240; CAA74636.1; -
 CC EMBL; X76939; CAA54258.1; -
 CC EMBL; Z99289; CAB16553.1; -
 CC HSSP; P02468; 1KLO.
 CC Gene; HGNC:6484; LAMA4.
 CC MIM; 600133; -
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR001791; Laminin_G.
 CC Pfam; PF00053; laminin_EGF; 6.
 CC Pfam; PF00054; laminin_G; 6.
 CC SMART; SM00180; EGF_Lam; 3.
 CC SMART; SM00282; LamG; 5.
 CC PROSITE; PS00022; EGF_1; UNKNOWN_1.
 CC PROSITE; PS01248; LAMININ_TYPE_EGF; 3.
 CC PROSITE; PS50025; LAM_G_DOMAIN; 5.
 CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 CC SIGNAL 1 24
 CC POTENTIAL.
 CC CHAIN 25 1816 LAMININ ALPHA-4 CHAIN.
 CC DOMAIN 82 131 LAMININ EGF-LIKE 1.
 CC DOMAIN 132 186 LAMININ EGF-LIKE 2.
 CC DOMAIN 187 240 LAMININ EGF-LIKE 3.
 CC DOMAIN 241 255 LAMININ EGF-LIKE 4 (INCOMPLETE).
 CC DOMAIN 256 825 DOMAIN II AND I.
 CC DOMAIN 826 1028 LAMININ G-LIKE 1.
 CC DOMAIN 1040 1220 LAMININ G-LIKE 2.
 CC DOMAIN 1227 1395 LAMININ G-LIKE 3.
 CC DOMAIN 1462 1633 LAMININ G-LIKE 4.
 CC DOMAIN 1640 1813 LAMININ G-LIKE 5.
 CC DOMAIN 313 396 COILED COIL (POTENTIAL).
 CC DOMAIN 466 521 COILED COIL (POTENTIAL).
 CC DOMAIN 574 607 COILED COIL (POTENTIAL).
 CC DOMAIN 655 717 COILED COIL (POTENTIAL).
 CC DOMAIN 770 799 COILED COIL (POTENTIAL).
 CC SITE 717 719 CELL ATTACHMENT SITE (POTENTIAL).
 CC BY SIMILARITY.
 CC DISULFID 82 91 BY SIMILARITY.
 CC DISULFID 84 98 BY SIMILARITY.
 CC DISULFID 101 110 BY SIMILARITY.
 CC DISULFID 113 129 BY SIMILARITY.
 CC DISULFID 132 146 BY SIMILARITY.
 CC DISULFID 134 155 BY SIMILARITY.
 CC DISULFID 157 166 BY SIMILARITY.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:42:08 : Search time 40 Seconds
(without alignments)
5738.412 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVGFPPYLLSAEVVK.....RPVLNLSLRAQLSHRLVLK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5864	100.0	1114	10 Q9Z58	Q9Z58 lycopersico
2	5070	86.5	1116	10 Q9ZRY7	Q9ZRY7 nictiana t
3	3588.5	61.4	775	10 Q9ZRX3	Q9ZRX3 petunia hyb
4	3532	60.2	1107	10 Q9LQV2	Q9LQV2 arabidopsis
5	1995.5	34.0	1133	10 Q82504	Q82504 arabidopsis
6	1984	33.8	513	10 Q9ZRX5	Q9ZRX5 arabidopsis
7	1621	27.6	1196	10 Q9SG02	Q9SG02 arabidopsis
8	1610	27.5	1196	10 Q9LKP0	Q9LKP0 arabidopsis
9	1088.5	18.2	1638	3 Q9C162	Q9C162 neurospora
10	1031	17.6	1484	3 Q8TGV4	Q8TGV4 diaporthe p
11	995	17.0	1483	3 Q96UL6	Q96UL6 diaporthe a
12	987	16.8	1122	3 Q8XIC2	Q8XIC2 phomopsis s
13	948	16.2	1215	3 Q14227	Q14227 schizosacch
14	932	15.9	1780	5 Q19285	Q19285 caenorhabdi
15	898	15.3	1579	5 Q9B856	Q9B856 caenorhabdi
16	896.5	15.3	1601	5 Q9NDH1	Q9NDH1 caenorhabdi

17	860.5	14.7	1632	5 Q93593	Q93593 caenorhabdi
18	720	12.3	1964	5 Q9ZG65	Q9ZG65 dictyosteli
19	703	12.0	1883	5 Q9ZG7	Q9ZG7 dictyosteli
20	515.5	8.8	1069	5 Q9XYP1	Q9XYP1 dictyosteli
21	456	7.8	107	10 Q9ZRX4	Q9ZRX4 triticum ae
22	441	7.5	953	3 Q01869	Q01869 phanerocha
23	385.5	6.6	1377	5 Q95WU3	Q95WU3 giardia lam
24	362	6.2	966	10 Q82190	Q82190 arabidopsis
25	355.5	6.1	1148	10 Q94IV7	Q94IV7 oryza sativ
26	350.5	6.0	929	10 Q82188	Q82188 arabidopsis
27	345.5	5.9	1222	10 Q94IV6	Q94IV6 oryza sativ
28	335.5	5.7	905	10 Q82189	Q82189 arabidopsis
29	297	5.1	1402	3 Q9Y7G6	Q9Y7G6 neurospora
30	145.5	2.5	634	4 Q96NH6	Q96NH6 homo sapien
31	140	2.4	2029	5 Q9VD07	Q9VD07 drosophila
32	138.5	2.4	1416	4 Q9BZF9	Q9BZF9 homo sapien
33	138.5	2.4	1449	6 Q9EG87	Q9EG87 bos taurus
34	134.5	2.3	549	3 Q9USN8	Q9USN8 schizosacch
35	134	2.3	1308	12 Q91DN2	Q91DN2 soil-borne
36	134	2.3	1816	12 Q91DN3	Q91DN3 soil-borne
37	133.5	2.3	2364	2 Q46342	Q46342 clostridium
38	133	2.3	2810	5 Q20456	Q20456 caenorhabdi
39	132	2.3	3072	12 Q92645	Q92645 clover yell
40	130.5	2.2	1091	16 Q9KAQ7	Q9KAQ7 bacillus ha
41	130	2.2	781	16 Q8XL45	Q8XL45 clostridium
42	129.5	2.2	849	10 Q9SA72	Q9SA72 arabidopsis
43	129	2.2	804	10 Q8W3J7	Q8W3J7 brassica ol
44	128.5	2.2	839	9 Q64076	Q64076 bacterioph
45	128.5	2.2	839	16 Q31945	Q31945 bacillus su

ALIGNMENTS

RESULT 1

Q9ZRS8 PRELIMINARY; PRT; 1114 AA.

AC Q9ZRS8
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48).
GN RDRP.

OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RUTGERS;

RX MEDLINE=9055198; PubMed=9836747;

RA Schiebel W., Fellissier T., Riedel L., Thalmair S., Schiebel R.,

RA Kempe D., Lottspeich F., Sanger H.L., Wassenegeger M.;

RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from

RT tomato.";

RL Plant Cell 10:2087-2102(1998).

DR EMBL; J10403; CAA71421.1; .

DR InterPro; IPR000504; RNA_rec_mot.

DR InterPro; IPR003955; RRM_2.

DR Pfam; PF00076; rrm; 1.

DR SMART; SM00362; RRM_3; 1.

KW Nucleotidyltransferase; RNA-directed RNA polymerase; Transferase.

SQ SEQUENCE 1114 AA; 126807 MW; B6D30ACD41DB37FF CRC64;

Query Match 100.0%; Score 5864; DB 10; Length 1114;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVGFPPYLLSAEVVKSFLEKTYGTGTCVCALEVKQSGSRAFAKVFADNISADK 60

DB 1 MGKTIQVGFPPYLLSAEVVKSFLEKTYGTGTCVCALEVKQSGSRAFAKVFADNISADK 60

```
QY 61 IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDMQDGIITLNFQCOISDDKFAVLGSTEVS 120
DB 61 IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDMQDGIITLNFQCOISDDKFAVLGSTEVS 120
QY 121 IOFGIGLKKFFFLSSGADYKQLQSYENITWQVYLHRPYGQNAQFLLIQLFGAPRIYKRL 180
DB 121 IOFGIGLKKFFFLSSGADYKQLQSYENITWQVYLHRPYGQNAQFLLIQLFGAPRIYKRL 180
QY 181 ENSCYSFFKETPDQWVRTDFFPSWIGLSSLCLOPRGRVRLPNFEESFFHYAERENNI 240
DB 181 ENSCYSFFKETPDQWVRTDFFPSWIGLSSLCLOPRGRVRLPNFEESFFHYAERENNI 240
QY 241 TLOFGTFFFSOKSALVPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVFFFLVDP 300
DB 241 TLOFGTFFFSOKSALVPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVFFFLVDP 300
QY 301 RRRNVACIEHALEKLYIKKCCYDVPVRLTEQYDGYLKGQPKSPSITLDDGLVYVRRV 360
DB 301 RRRNVACIEHALEKLYIKKCCYDVPVRLTEQYDGYLKGQPKSPSITLDDGLVYVRRV 360
QY 361 LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR 420
DB 361 LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR 420
QY 421 TNIVYERILSTLRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ 480
DB 421 TNIVYERILSTLRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ 480
QY 481 IKNVAKYAARLGQSGSSRETLSVLRIHEIEVTPDVKHGTSYVFSOGIGKISGDFAHRYA 540
DB 481 IKNVAKYAARLGQSGSSRETLSVLRIHEIEVTPDVKHGTSYVFSOGIGKISGDFAHRYA 540
QY 541 SKCGLQYTPSAFQIRYGGYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC 600
DB 541 SKCGLQYTPSAFQIRYGGYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC 600
QY 601 YLNQLTLTLSTLGKQDEVLEQKQKAVDQDAILHDSLKAQALELMSGENTNLIKAM 660
DB 601 YLNQLTLTLSTLGKQDEVLEQKQKAVDQDAILHDSLKAQALELMSGENTNLIKAM 660
QY 661 LNCGYKPDABEFLSMILQTPRASKLLDLRTSRIFIPNGRTMGCGLDESRTLYGQVFFVQ 720
DB 661 LNCGYKPDABEFLSMILQTPRASKLLDLRTSRIFIPNGRTMGCGLDESRTLYGQVFFVQ 720
QY 721 FTGAGHGEFSDDLHPFNNSRSTNSNFIKNGNVVAKNCPCLHPGDIIRYLKAVNVRLAHMV 780
DB 721 FTGAGHGEFSDDLHPFNNSRSTNSNFIKNGNVVAKNCPCLHPGDIIRYLKAVNVRLAHMV 780
QY 781 DCVVEPKGKRPHNECGSDLDGDIYFVCWDQDMIPPROVQPMEXPPAPSIQLDHDTV 840
DB 781 DCVVEPKGKRPHNECGSDLDGDIYFVCWDQDMIPPROVQPMEXPPAPSIQLDHDTV 840
QY 841 BEVEEYETNIVNDSLGIIANAHVVFADRPDMAMSDPCCKLAELPSIAVDFFPKTVPAE 900
DB 841 BEVEEYETNIVNDSLGIIANAHVVFADRPDMAMSDPCCKLAELPSIAVDFFPKTVPAE 900
QY 901 IPSOLRKEYPDFMDKPDKTSYISERVIGKLFKRVKDKAPQASSIAFTETRDVARRSYDAD 960
DB 901 IPSOLRKEYPDFMDKPDKTSYISERVIGKLFKRVKDKAPQASSIAFTETRDVARRSYDAD 960
QY 961 MEVDGFEDYIDEAPDYKTEYDNKLGNDLMDYIGIKTEAEILSGGIMKASKTFFDRKKDAEAI 1020
DB 961 MEVDGFEDYIDEAPDYKTEYDNKLGNDLMDYIGIKTEAEILSGGIMKASKTFFDRKKDAEAI 1020
QY 1021 SVAVRALRKARAWFKRRNDIDMDLPKASAWYHVYHTYWGVCYNGLKRAHFISPPWCY 1080
DB 1021 SVAVRALRKARAWFKRRNDIDMDLPKASAWYHVYHTYWGVCYNGLKRAHFISPPWCY 1080
QY 1081 YDQLIQIKKDKARNRPNVNLSSSLRAQLSHRLVLK 1114
DB 1081 YDQLIQIKKDKARNRPNVNLSSSLRAQLSHRLVLK 1114
```

```
RESULT 2
Q9ZRY7 PRELIMINARY; PRT: 1116 AA.
AC Q9ZRY7;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RNA-directed RNA polymerase.
GN RDRP.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CV. PETIT HAVANA SRI; TISSUE=LEAF;
RA Schiebel W., Pellissier T., Riedel L., Thalmeir S., Schiebel R.,
RA Kempe D., Lottspeich F., Saenger H.L., Wassenecker M.;
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT Tomato.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
KW EMBL; AJ011576; CAA09697.1; -.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1116 AA; 127553 MW; 8D9889DA826D31DB CRC64;

Query Match      86.5%; Score 5070; DB 10; Length 1116;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 959; Conservative 67; Mismatches 87; Indels 2; Gaps 1;

QY 1 MCKTIQVFGFPYLLSAEVVKSLEYKTYGTVCALELVKQSGKGRAFRAKQVADMISADK 60
DB 1 MCKTIQVFGFPYLLSAEAIVKKELENHTNGTYVALELVKQSGKGRAFRAKQVANKRAEF 60
QY 61 IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDMQDGIITLNFQCOISDDKFAVLGSTEVS 120
DB 61 IIDLASKGYIYGPYLKAWEMKTDIVOPRTYVHRMDGVTLNFQCOISDDKFAVLGSKDVS 120
QY 121 IOFGIGLKKFFFLSSGADYKQLQSYENITWQVYLHRPYGQNAQFLLIQLFGAPRIYKRL 180
DB 121 IKFGIGLKKIYFFLHASADYKQLQSYENITWQVYLHRPYGQNAQFLLIQLFGAPRIYKRL 180
QY 181 ENSCYSFFKETPDQWVRTDFFPSWIGLSSLCLOPRGRVRLPNFEESFFHYAERENNI 240
DB 181 ENSCYSFFKETPDQWVRTDFFPSWIGLSSLCLEFRNGVOLPNFSEFFYKESMNOF 240
QY 241 TLOFGTFFFSOKSALVPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVFFFLVDP 300
DB 241 ILQGTFTFSOKLALVPIVHLPEGLPELPYKILFKISSLIQHGCCFGLALNPNFQLVDP 300
QY 301 RRRNVACIEHALEKLYIKKCCYDVPVRLTEQYDGYLKGQPKSPSITLDDGLVYVRRV 360
DB 301 RRRNVACIEHALEKLYIKKCCYDVPVRLTEQYDGYLKGQPKSPSITLDDGLVYVRRV 360
QY 361 LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR 420
DB 361 VVTPCKYVFCGPEVNVNRVLRNYSEDINFLRVSFVDEWEKIHSTDLLPRASTGNGTR 420
QY 421 TNIVYERILSTLRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ 480
DB 421 TDIVYERILSTLRNGFTIGDKREEELAFSSQDRDNVSMFASRPGLTANDIRAWMGDFRQ 480
QY 481 IKNVAKYAARLGQSGSSRETLSVLRIHEIEVTPDV--KVHGTSYVFSOGIGKISGDFAH 538
DB 481 IKNVAKYAARLGQSGSSRETLSVGRHEVEIPDVVACSLHGTNYIFSDGIGKISADFAHR 540
QY 539 VASKGLQYTPSAFQIRYGGYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQ 598
DB 541 VTIKGGLQYTPSSFTIRYGGYGVGVAVDPYSSMKLSLRKSMKSYESDNIKLDVLGWSKYQ 600
QY 599 PCYLNRLQTLTLSTLGKQDEVLEQKQKAVDQDAILHDSLKAQALELMSGENTNLIK 658
DB 601 PCYLNRLQTLTLSTLGVIDVLEQKQNEAVDQDAILHDSLKAQALELMSGENTNLIK 660
```



```
QY 659 AMLNCGYKPDAPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYQGVF 718
Db 661 EMLNCGYMPDAEPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYQGVF 720
QY 719 VQFTGAGHGFSDDLHPFNNSRNSNFILKGNVVAKNCPCLHPGDIRVLKAVNRALHH 778
Db 721 VQFSGARRQFESHSFNFSGSANCDFILKGNVVAKNCPCLHPGDIRVLKAVNRALHH 780
QY 779 MVDGWFFQKRPHPNECSGSLDGDIIYFVWPDQDMIPPRQVQPMPEYPPAPSIQLDHDV 838
Db 781 MVDGWFFQKRPHPNECSGSLDGDIIYFVWPDQDMIPPRQVQPMPEYPPAPSIQLDHDV 840
QY 839 TIEVEVEFYNYINDSLGIITIANAHVVFADREPDMAMSDPCKLAELFSIAVDFPKTGP 898
Db 841 TIEVEVEFYNYINDSLGIITIANAHVVFADREPDMAMSDPCKLAELFSIAVDFPKTGP 900
QY 899 AEIPLSQRPRKPEYDPMKPKPTYPSEKVIKLFQKVKNTQPCASSIATFTRVARKSYD 958
Db 901 AEIPLSQRPRKPEYDPMKPKPTYPSEKVIKLFQKVKNTQPCASSIATFTRVARKSYD 960
QY 959 ADMEVDGFEDYIDAFYKTEYDNKGLNMDYIGIKTEAEILSGGIMKASKTFDRRKDAE 1018
Db 961 SDMIVDGFEDYIDAFYKTEYDNKGLNMDYIGIKTEAEILSGGIMKASKTFDRRKDAE 1020
QY 1019 AISVAVRALRKEARAWFKRRNDIDMLPKASAWHYHYHTYGYCYNQGLKRAHFISFPW 1078
Db 1021 AIGVAVCLRKEARAWFKRRNDIDMLPKASAWHYHYHTYGYCYNQGLKRAHFISFPW 1080
QY 1079 CVYDQLIOIKDKARNRPNVLSLRAQLSHRLVL 1113
Db 1081 CVYDQLIOIKDKARNRPNVLSLRAQLSHRLVL 1115

RESULT 3
Q92RX3 ID Q92RX3 PRELIMINARY; PRT; 775 AA.
AC Q92RX3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RNA-directed RNA polymerase (fragment).
GN RDRP.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Schiebel W., Pelissier T., Riedel L., Thalmair S., Schiebel R.,
RA Kempe D., Lottspeich F., Saenger H.L., Wassenegger M.;
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT Tomato."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011979; CAA09896.1; -
KW RNA-directed RNA polymerase.
FT NON_TER 1
FT NON_TER 775
SQ SEQUENCE 775 AA; 88778 MW; DE43FE50C6A45B93 CRC64;

Query Match 61.4%; Score 3598.5; DB 10; Length 775;
Best Local Similarity 87.0%; Pred. No. 7.7e-251;
Matches 675; Conservative 47; Mismatches 53; Indels 1; Gaps 1;

QY 307 CIEHALEKLYIKCCYDPPRWLTQEQDGYLKGQPKPKSPSITLDGLVVRVRLVTPCK 366
Db 1 CIEHALEKLYIKCCYDPPRWLTQEQDGYLKGQPKPKSPSITLDGLVVRVRLVTPCK 60
QY 367 VYFGPEVNVNRYLRNYSIDINFLRVSVDEWEKLYSTDLLPKASTGSGVRTNIYER 426
Db 61 VYFLGPEVNVNRYLRNYSIDINFLRVSVDEWEKLYSTDLLPKASTGSGVRTNIYER 119
```

```
QY 427 ILSTLRKGEVIGDKFEFLAFSSQLRDRNSVVMFASRPGLTANDIRAMWGDFSQIKNVAK 486
Db 120 ILSTLQNLVIGAKKFEFLAFSSQLRDRNSVVMFASRPGLTAYDIRTWMGDFQOIRNVAK 179
QY 487 YAARLQGSQSSRETLSVLRHETEVIPDVKHGTSVFSDGIGKISGDFAHFVASKCGLQ 546
Db 180 YAARLQGSQSSRETLSVGRHEIKYIPDVEVHGTNVFSDGIGKISADFAHRVAKCGLQ 239
QY 547 YTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSYEDSNILKDLVLGWSKYQPCYLNRL 606
Db 240 HTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSYEDSNILKDLVLGWSKYQPCYLNRL 299
QY 607 ITLLSTLGVKDEVLQKQKAEVNDQDAIILHDSLKQAELALMSPGENTILKXMLNCGYK 666
Db 300 VTLMSTLGVKDEVLQKQKAEVNDQDAIILHDSLKQAELALMSPGENTILKXMLNCGYK 359
QY 667 PDAEPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYQGVFQFTGAGH 726
Db 360 PDAEPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYQGVFQFTGAGH 419
QY 727 GFESDDLHPFNNSRNSNFILKGNVVAKNCPCLHPGDIRVLKAVNRALHHMVDGVPP 786
Db 420 RQFYEEHSFNFSGSANYNFILKGNVVAKNCPCLHPGDIRVLKAVNDVPALHHMVDGVPP 479
QY 787 QKGRPHPNNECSGSLDGDIIYFVWPDQDMIPPRQVQPMPEYPPAPSITLDHDVITIEVEY 846
Db 480 QKGRPHPNNECSGSLDGDIIYFVWPDQDMIPPRQVQPMPEYPPAPSITLDHDVITIEVEY 539
QY 847 FTYNYVNDSLGIITIANAHVVFADREPDMAMSDPCKLAELFSIAVDFPKTGPVPAEIPSQL 906
Db 540 FTYNYVNDSLGIITIANAHVVFADREPDMAMSDPCKLAELFSIAVDFPKTGPVPAEIPSQL 599
QY 907 PREYDPMKPKPTYPSEKVIKLFQKVKNTQPCASSIATFTRVARKSYDADMEVDGF 966
Db 600 PREYDPMKPKPTYPSEKVIKLFQKVKNTQPCASSIATFTRVARKSYDADMEVDGF 659
QY 967 EDYIDEAFDYKTEYDNKGLNMDYIGIKTEAEILSGGIMKASKTFDRRKDAEISAVARA 1026
Db 660 EDYIDEAFDYKTEYDNKGLNMDYIGIKTEAEILSGGIMKASKTFDRRKDAEISAVARA 719
QY 1027 LRKEARAWFKRRNDIDMLPKASAWHYHYHTYGYCYNQGLKRAHFISFPWCYVD 1082
Db 720 LRKEARAWFKRRNDIDMLPKASAWHYHYHTYGYCYNQGLKRAHFISFPWCYVD 775

RESULT 4
Q9LOV2 ID Q9LOV2 PRELIMINARY; PRT; 1107 AA.
AC Q9LOV2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F1086.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,
RA Williams S., Altairi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA Hansen N.F., Huizar L., Kremenetskaia I., Ienz C., Li J., Liu S.,
RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome
RT I."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
```



```
[3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN
[4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
  Khan S., Kim C., Altati H., Bei B., Chin C., Chou J., Choi E.,
  Conn L., Conway A., Gonzalez A., Hansen N., Howang B., Koo T., Lam B.,
  Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,
  Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
  Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
  Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006917; AAF79241.1;
SQ SEQUENCE 1107 AA; 126214 MW; F8FF97691FF82187 CRC64;

Query Match 60.2%; Score 3532; DB 10; Length 1107;
Best Local Similarity 62.4%; Pred No. 8.6e-246;
Matches 692; Conservative 160; Mismatches 225; Indels 32; Gaps 12;

QY 1 MGKTIQVGFPGYLLSAEAVKSWFLKTYGTGVCALVEYKQ-SKGSRAFAKVPADNISAD 59
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MGKTIQVGFPGYVSAEAVKSWFLKTYGTGVCALVEYKQ-SKGSRAFAKVPADNISAD 59
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 60 KIITLANNRLYFGSSYLKANEMKTDIV-QLRAYVDQMDGITLNGCOISDDKEAVLGST- 117
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 LIITAAERUYIGRSYLKAFEEVDIVPKPRASLHTISGLKMFPGCOVTKKFLTLWSAQ 120
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 118 EVSIQFGLGKFFFFSSSADYKLOLSYENIWOVYLHRPYGONAFLLIQLFGAPRIY 177
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 DVCVSFGIGMKLHFSWYKQYRLELSYENIWOIDLHSPQGRSFLYIVQIGAPKIF 180
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 178 KELEN-----SCYSFPEKPEDDOWRTDF-PFSWIGLSSSLCLOPRGVRPLNPEESF 230
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
DB 181 EKEDQINLFLGIMDFYSDGSDEQMIERTDFTSSSCIGQSTAFCELPVHLNVPDFRNF 240
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|

QY 231 PHYAE-RENNITLTQFTFFYSQKSAVNPVQPPGEGISIPYKILFKLSSIVQHGCIQPGA 289
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
DB 241 ANYAEHRASSFLIESG-SYSSNNANTLVVPPPPGSLPFEILFKLTLVQNACLSGPA 299
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|

QY 290 LNVFFRLVDPRRNAVACIEHALKLYIKECCYDPVWLTQYDGYLKGQPKPSPIIT 349
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
DB 300 LDLDYFLLNKKYDRALIDHCKLEKFLGECCEYEPAAHLRDEYKWKISGKLPLSPSTIS 359
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|

QY 350 LDDGLVYVRRVLTVPCKYFCGPPVNVSNRVLNRYSDIDNFRVSPVDEWEKLYSTDL 409
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
DB 360 LDDGLVYMRVQVTPARVYFSGPVPVNVSNRVLRYSKYINNFLEVSFVDEDELEKYSMDL 419
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|

QY 410 LPKASTGSGVRTNIYERILSTRKGFVIGDKKFEFLAFSSQLRDSNVMWFASRPGILTAN 469
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
DB 420 SPRGSQTR--RTKLYDRIYSVLRDGVIGDKKFEFLAFSSQLRDSNVMWFASRPGILTAN 477
  ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|

QY 470 DIRAMGDGFSQIKNVAKYAAALGSGSSRETLVSRHEIEVTPDVKV--HGTSYFSDG 527
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 478 HIRAMWGDGFIHNVAKYAAALGSGSSRETLVSRHEIEVTPDVKV--HGTSYFSDG 537
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 528 ICKTSGDFAHFVASKGL-QYTPSAFQIRYGGYGVYGVDPDSSMKLSLRKSMKSYESDN 586
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 538 ICKTSGDFAHFVASKGL-QYTPSAFQIRYGGYGVYGVDPDSSMKLSLRKSMKSYESDN 597
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 587 IKLDVLGWSKYQPCYLNRLQTLITLSTLVGKDEVLVEQKQKAVDQDLDAILHDSLKQAPALE 646
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 598 TKLDVLGWSKYQPCYLNRLQTLITLSTLVGKDEVLVEQKQKAVDQDLDAILHDSLKQAPALE 657
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 647 LMSPGENTINILKAMNGCYKDPDAPFFLSMMLQTPFRASKLLDLRTSRIFIPNGRTMMGCL 706
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 658 LMAFGENTINILKALILCYKDPDAPFFLSMMLQTPFRASKLLDLRTSRIFIPNGRTMMGCL 717
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 707 DESTLEYQGVVQVFTGAGHGEFSDDLHPFNNSSTNSNFIKGNVYVAKNPCLHPGDIR 766
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 718 DETRTLEYQGVVQVFTGAGHGEFSDDLHPFNNSSTNSNFIKGNVYVAKNPCLHPGDVR 762
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

DR EMBL; AF080120; AAC35535.1; -
DR EMBL; AL049876; CAB43048.1; -
DR EMBL; AL161531; CAB81214.1; -
KW Nucleotide transferase, RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 1133 AA; 129323 MW; 32B72C4E429B20B9 CRC64;

Query Match 34.0%; Score 1995.5; DB 10; Length 1133;
Best Local Similarity 40.1%; Pred. No. 7e-135;
Matches 459; Conservative 214; Mismatches 389; Indels 83; Gaps 31;

QY 4 TTQVGFYLLSAEVVSEFLEKYTGCTVCALEVQKSGG--SRAFAKVQFAD-NISADK 60
DB 11 TTKISNVPTIIVADELLRFLHGEDIVFALEIPTTRDNMKPRDFARVQTTLEVKRA 70
QY 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVD--QMDGITLNFQGISDCK-FAVLGS 116
DB 71 QLLSSQSLLFKTHNLRLEAYDIIIP--RPVDPKRLLDDIVLTVPFESDEKRFCALEK 128
QY 117 TEVSTQFGIGLKK---FFFFSSGADYKQLQSLNINWQVHLRHPYQONAF--LLIQLF 171
DB 129 WD-GVRCWILTEKRVRVEFWWESGDC-YKIEVRFEDIITLSCCVNGDSASEIDAFLLKL 186
QY 172 GAPRIYKRL-----ENSCYSFFKETPDQDQWVKTDFPPS-WIGLSSSLCLOFRGVR 222
DB 187 YGPKYKRVTHIAKFKSDRYCKEDFDPMWITTFDFSGSKSIGTSCFLEWHNGST 246
QY 223 LPNFESEFFHAERNNITLOTGTFVFSQKSAALVPNVQPP-EGISIPYKILFKISSIVQ 281
DB 247 MLDIFSGLPYREDTSLSTYVDGKTF--ASAAQIVPLNAAAILGLEFFYEILFOINLVH 304
QY 282 HGCIPGALNVYFRLVDPRRNV--ACIEHA---LEKLYIKECVDPVRWLTQEVYDG 335
DB 305 -----AQISLFAASDMELIKILRGMSLETALVILKLLHQSSICVDVFFVKTQMS 357
QY 336 YLK--GROPPKSPSTLDDGLVYVRVRLVTPCKVFCGPEVNVSNRNLRYSEIDNPLR 393
DB 358 VYKMKHSPASAYKRLTEQINMQORAVTPTSKYLLGPELETANYVVKNAEHSVDFMR 417
QY 394 VSFVDEWEKLYSTDLLPKASTGSGV---RNIYERILSTLRKGFVIGDKKFEFLARSS 450
DB 418 VTFVEDSKJPANALSVNSKEGVFVKPSRTNINRVLISILGEGITVGPKRFEFLAFAS 477
QY 451 QLDRNSVMEFASRPGLTANDIRAMGDFSQIKNVAKYAALQGSGSRETLSVLRIEIE 510
DB 478 QLGRNSVMEFASNEKVAEDIREWNGCFKIRISIKCAARMGQLFSASRQLIIVRAQDVE 537
QY 511 VTPDVKV--HGTSYVFSOGIKISGDFAHRVASKGCGLYTPSAFQIRYGGYKGVGVDDP 568
DB 538 QIPDIEVTDGADYCFDSDGIGKISLAFKQVAKQKGLSHVPSAFQIRYGGYKGVIAVDRS 597
QY 569 SSKLSLRKSKYSBDSNKLKDLGWSKYQPCYNLRQLITLLSTLGKDEVLEQKQEA 628
DB 598 SFRKLSLRDSMLKFDSSNRLNVRWTSMPCLNREIICLLSLTGLTEDANFEAMQAVHL 657
QY 629 DQLDAILHDSLKAQPALELMSGENT-NILKAMLNCGYKPDAPFPLSMMLQTFRASKLLD 687
DB 658 SMLGNMLEDRAALNVQLKLS-GENSKNLLVKMLLQGVASSEYPSLMLRVHSESQSL 716
QY 688 LRTSRIFIPNGRTMMGCLDSRTLEYGVQVQFTGAGHGEFSDLLHPFNNSRSTNSFI 747
DB 717 LKSRCLILVPGRIILIGCDMEGLLEYGVYVRVTLTKAEKLSRQDSYFRKIDETS--V 774
QY 748 LKGNVYVAKNCPHLPDGIIRLVKAVNRLHH---MVDVYVFPQKGRPHNECSGSDLD 803
DB 775 VIGKVVYTKNCPHLPDGIIRVLA--IYEVHFEKGYLDCIIFPKQGERPHNECSGGDLD 832
QY 804 GDIVFVCHDQMIIPRQVQMEYPPAPSIQLDHDVTIEVEEYETNYINVDLSLIANAH 863
DB 833 GQFFVSVDEKILISEMPPMDYAGSRPLMDHDVTEEHKFFVYIMISITLGVISTAH 892
QY 864 VVADREPDMAMSPCKKLAEFLSIADVFPKTVGTPAEIPSPKPEYDFDMKDKPYSI 923
DB 893 LVHADROPKARSQKCLELANLHSPRAVDFAKTGAPAEYPAALKPREPDPFLERKEPTYI 952

QY 924 SERVIGKLFKRVK----DKAPOASSIATFTEDVARRSYDADMEVDGDPEDYIDEAFDYKTE 979
DB 953 SESVFGKLYRAVKSLSLAQRPEASEDTV-----AYDVTLEAGFESTIETAKAHRDM 1005
QY 980 YDNKLGMLDYYGIGKTEAEILSGIGMKASKTF---DRRKDA--EASIVAVRALRKEARA 1033
DB 1006 YGKLTSLMIYGAANEIEILT-GILTKEMYLARDNRRYGMKDRITLSVKDLHKEAMG 1064
QY 1034 WFKRNDIDMLPK-ASAWTHVYHPTYWCYNOGLKRAHFISFPWCYVQQLTIQKKDKA 1092
DB 1065 WFEKSEDEQCKKLASAWYVYTNP-----NHRDEKLTFLSPFWIVGVLDLKAENA 1118
QY 1093 RNRPV 1097
DB 1119 QRQSV 1123

RESULT 6
Q92RX5 PRELIMINARY; PRT; 513 AA.
AC Q92RX5;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE RNA-directed RNA polymerase (fragment).
GN RDRP.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA; TISSUE=LEAF;
RA Schiebel W., Pelissier T., Riedel L., Thalmair S., Schiebel R.,
RA Kempe D., Lottspeich F., Saenger H.L., Wassenecker M.;
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT Tomato";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011977; CAA09894.1; -
KW RNA-directed RNA polymerase.
FT NON_TER 1 1
FT NON_TER 513 513
SQ SEQUENCE 513 AA; 57802 MW; 924FDA2B70EDA0E CRC64;

Query Match 33.8%; Score 1984; DB 10; Length 513;
Best Local Similarity 72.9%; Pred. No. 1.3e-134;
Matches 385; Conservative 50; Mismatches 75; Indels 18; Gaps 4;

QY 471 IRAMWGDFSOIKNVAKYAALQGSGSRETLSVLRIEIEVDPVKV--HGTSYVFSDDGI 528
DB 1 IRAMWGDFDHIRNVAKYAALQGSGSRETINVRSDIEIVDFEIIISLGTIFYVSDGI 60
QY 529 GKISGDFAHRVASKGL-QYTPSAFQIRYGGYKGVGVDDPSSMKLSLRKSKYSBDSNI 587
DB 61 GKISAEFARVARKCGLTEFSPSAFQIRYGGYKGVAVDPNSSKLSLRKSKYSBSEST 120
QY 588 KLDVLGWSKYQPCYNLRQLITLLSTLGKDEVLEQKQEAVDQDLAILHDSLKAQEALE 647
DB 121 KLDVLWASKYQPCYNLRQLITLLSTLGVTDSVFEKQREVVDRDLAILHDSLKAQEALE 180
QY 648 MSPGENTNILKAMLNCGYKPDAPFPLSMMLQTFRASKLLDRTSRIFIPNGRTMMGCLD 707
DB 181 MAPGENTNILKALITCGYKPDAPFPLSMMLQNFASKLLELRKTRIFISGGRSMGCLD 240
QY 708 ESRILEYGVQVQFTGAGHGEFSDLLHPFNNSRSTNSFIKGNVYVAKNCPHLPDGI 767
DB 241 ETRTLEYGVVWQ-----YSDMPRP-----GRRFITGFPVVVAKNCPHLSGDVR 285
QY 768 LKAVNVALHMHVDCVVFPQKGRPHNECSGSDLDGDIYFVCWDQDMIPPRQVQPMYEP 827
DB 286 LQAVNVPALNHMWDVVFPQKGLRPHNECSGSDLDGDIYFVCWDQELVFPRTSEPMYDT 345

```
Qy 828 PAPSQILDHVTIEVEYFTNYIVNDSGLIIAHNVFADREPDMAKSDPCKLAELFS 887
Db 346 PEPTQILDHVTIEVEYFANYIVNDSGLIIAHNAHAFADKEPLKAFSPDCIELAKFS 405
Qy 888 IAVDFPKTGPAPKPEYDFDKDKTYSISERVIGKLFKRVKAPQASSIAT 947
Db 406 TAVDFPKTGAAPVTPQHLHYKEYDFDEMEKDPKPYESKNVIGKLFREVKERAPPLISIKS 465
Qy 948 FTRDVARASYADMEVGFEDYIDEADYKTEYDNKLGNLMDYGIKT 995
Db 466 FTLDVASKSYDKMEVDGFEYVDEAFYQKANYDFKLGNLMDYGIKT 513

RESULT 7
Q9SG02 PRELIMINARY; PRT; 1196 AA.
ID Q9SG02
AC Q9SG02;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Putative RNA-directed RNA polymerase.
GN TIG12.20 OR T9C5.95
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome 1 BAC TIG12 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Rieger H.W., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,
RA Meves H.W., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20306668; PubMed=10850495;
RA Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C.,
RA Morel J.-B., Jouette D., Lacombe A.-M., Nikic S., Picault N.,
RA Renoue K., Sanial M., Vo T.-A., Vaucheret H.;
RT "Arabidopsis SGS2 and SGS3 Genes are Required for Posttranscriptional
RT Gene Silencing and Natural Virus Resistance.";
RL Cell 101:533-542(2000)
DR EMBL; AC012329; AAG52184.1; -
DR EMBL; AL132964; CAB71285.1; -
DR EMBL; AF239718; AAF71959.1; -
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136927 MW; 812DEB9CEDC82C8F CRC64;

Query Match 27.68; Score 1621; DB 10; Length 1196;
Best Local Similarity 34.68; Pred. No. 8.2e-108;
Matches 431; Conservative 191; Mismatches 414; Indels 210; Gaps 42;

Qy 5 IQVFGFPVLLSAEVKSPLEKYTGVTGTCALAEVQSKGGSRAFAKVOFADNI-SADKIIT 63
Db 15 VSIIGFGESTAKQLTDYLEDEVGIVWCRUKLTWTPGSPYFNEIADTSNIPIDEYK 74
Qy 64 LANNELYGSSYLKAWEMKTDIVLRAYVD-----QMDG 97
Db 75 VEPFAVHFVAFSPASGRAMDAGQCNLLIDGPKLVSLGPKNPYSLNQRRTTTPYKLAG 134
Qy 98 ITLNFQCOISDDKFAVLGSTEVSIOFGI----GLKKFFELSSG-----SADY 141
Db 135 ITLEIGTLVSRDDFFVSVRAE-GVDFLVDPDNCTCKFRKSTAFSKDAYNHVINC 193
```

```
Qy 142 KLQLSYENLQWVLRHPYGGNAQFLLI-QLFGAPRIYKRLSENCYSFFKTEP-----DD 194
Db 194 KLELLVRDIQTV---ROYKTLHGVLQLASSPRVWYTADD--DIYTVFGDLLDDDD 248
Qy 195 QWVTTDFPPSSIGLSSSLCLOFRGRVLPNFEESEFFHYAERENNITLOTGTFF---V 250
Db 249 PWIRTTDF--TVG-AIGRCHSVRLIS-PRYENK-----LRTALDIYFMRRV 292
Qy 251 SOKSALVPN-----VOPPEGISIPYKILPKRKLSSLVQHGCIPGALNV 292
Db 293 QEERVWPPRIRNEPCFGEVSDHFFCIHKGES--FEIMFLVNSVLHGVNFQFLTE 350
Qy 293 YFFRLV--DPRRENACIEHALEKLYIIRECCYD--PVRWLTEQYDGYLK-----GRQ 341
Db 351 RPFLLRNQPKDQVNIASLAKL-----CTYKREV-----FDAYKRLKLVQEWIQKN 395
Qy 342 PKASPSITLDDGLVYVRRVLVTPCKVYFGPEVNVSNRVLRYNSEDIDNLFVSVFDEEM 401
Db 396 PKLLGSHSESEDISEIRLVITPTRAICLPPEVELSNRVLRYKAVAEFLRVTFMDES 455
Qy 402 EKLYSTDL-----LPKASTGSGV--RTNIYERILTLRKGFGVIGDKKTEFLAFSSQL 453
Db 456 QTINSNVLVYFVAPIVKDLTSSFSQKTVYFRKVSILTDGFKLCGRKYSFLAFSANQR 515
Qy 454 DNSVWFAFRPGLTANDIRAWMGDFSQIKNVAKYARLQSGSSRETLSVLRIHEIV-I 512
Db 516 DRSAWFAEDGKTRVSDIKTWGKFKD-KNVAKCAARMGLCFSSYATYDVMPHEVDTEV 574
Qy 513 PDVKVHGTSYVFSDDGIGIKSGDFAHRVASKGL--OYTPSAFOIRYRGYGVGVDPDSS 570
Db 575 PDIERNG--YVSDGIGITIPDLADEVMEKLLDVHVSFCAYOIRYAGFGVVARWPSKS 632
Qy 571 --MKLSLRKSMKSYESDNTKLDVLGSKYOPCVNLNQLITLTLSTLGVKDEVLEQKQEA 628
Db 633 DGIRLALRDSMKFFSKFHITLICSWTFRQPGFLNQIITLLSVLGPDPDEIWMQESML 692
Qy 629 DQIDALIHDSLKAQEALEMLSPGENTNLIKAMLCNOCYKPDABEFLSMQLQTFRASKL 688
Db 693 YKLNRLDDDFDAFEVL-TASCAEQNTAAILMSAGFKPKTEPHLRGMLSSVRIAQLWGL 751
Qy 689 RTRSRIFPNGRMTMGLCDESRTLEYGOVFOPT-----GAGHGEFSDDLHPFN 737
Db 752 REKSRIFTVSGRWLGLDEAGILEHGQCQFIQVSKPSIENCFSKHGSREKTKTDLE--- 808
Qy 738 NSRSTNSNFIKGNVVVAKNPCLHPGDIVLKVANNRHALHHMVDVCFVPOKGRPHNEC 797
Db 809 -----VVKGYIAIKNPCLHPGDVRIEADVLPQLHHMYDCLIFKQKGRPHNEA 859
Qy 798 SGSDLDGDIYFVCWQDMIPP-RQVOP-MEYPPAPSIQLDHDVTIEVEYFTNYIVNDS 855
Db 860 SGSDLDGDIYFVAVDQKLIIPPNRKSYPAHMDAAEKSILGRAVNHQDIIDFFARNLANEQ 919
Qy 856 LGIIANAHVVFADREPDMAKSDPCKLAELFSIADVFPKTVGPAPKPEYDFDEMD 915
Db 920 LGTICNAHVHADREYSGAMDEECLLAEALAVDFPKTKIVSMPPFHLKALYIPDFMG 979
Qy 916 KPDKTSYISERVICKLFRVKV---DKAPOASSIATFTRDVARASYADMEVGFEDYIDE 972
Db 980 KEDYQIVKSNKILGRLYRRVKEVYDEDAEASS--EESTDPSAIPYDAVLEIPGFEDLPE 1037
Qy 973 AFYKTEYDNKLGNDYGIKTEAEILSGIMKASKTDRRKD--AEAISVAVRALKE 1030
Db 1038 AWGHKCLYDQGLIGLLGQYKQVEEIVTGHVMSMPKYTKKQKELKRLKHSYLSKKE 1097
Qy 1031 ARAWFKRRNDIDMWLP-----KASAWYHVHTYHTYGCYNOGLK-----1069
Db 1098 FRKVF-----ETIDPHENLSEEEKNILEKASAWYHVHTYHPE-VWKKSLELQDPDES 1150
Qy 1070 -RAHFSFPWCYVDQLTIQK-----KDKARNRVLNLSLRAQ 1106
Db 1151 SHAAMLSFAMIAADYLARIKIRSGMSIDSA--KPVDSLAKFLAQ 1194
```

```
RESULT 8
Q9LKP0 PRELIMINARY; PRT; 1196 AA.
AC Q9LKP0;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE RNA-dependent RNA polymerase.
GN SDE1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20306669; PubMed=10850496;
RA Dalway T., Hamilton A., Rudd S., Angell S., Baulcombe D.C.;
RT "An RNA-dependent RNA polymerase gene in Arabidopsis is required for
RT posttranscriptional gene silencing mediated by a transgene but not by
RT a virus.";
RL Cell 101:543-553(2000).
KW EMBL; AF268093; AAF74208.1; -.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136747 MW; C1B1961CDC11F708 CRC64;

Query Match 27.5%; Score 1610; DB 10; Length 1196;
Best Local Similarity 34.6%; Pred. No. 5.1e-107;
Matches 431; Conservative 188; Mismatches 417; Indels 210; Gaps 42;

QY 5 IQVFGPYLLSAEVAKSFLEKYTGTCVCALEVQKSGGSAFAKYOAFADNI-SADKIIT 63
DB 15 VSIGFGESITAKLTQDLEDEGVWRCRLKTSWTPPGSYNFEIADTSNIPSIDEYKK 74
QY 64 LANNRYFGSSYLKANEMKTDIVQLRAYD-----OMDG 97
DB 75 VEPHAFVHFAVFSAGRAMDAOCNLLDQPLKVS LGPKNPSYLNQRRTVTPKLAG 134
QY 98 ITLNFQCOISDDKFAVLGSGTEVSTQFGI-----GLKQFFFLSSG-----SADY 141
DB 135 ITLEIGLVSRDDEFVSWRAE-GVDFLVDPFDNTCKFCPRKSTAFSKDAMHAVINCDY 193
QY 142 KLOLSYENIQOVULHRYPGONAFLLI-QLFGAPRIYKRLNESCYSFFKTP-----DD 194
DB 194 KLELLVRDIQT-----ROYKTHGFVLVLQLASSPRVWYRTADD--DIYDFVPGDLLDDDD 248
QY 195 QWVRTDFPPSWIGLSSSLCLOPRGRVRLNFEESFFHYAERENITLQGFTEFF-----V 250
DB 249 PWIRTTDF--TQVG-ALGRCHSYRLVIS-PRYENK-----LRTALDYFMRVY 292
QY 251 SQKSALVPN-----VQPPGEGISIPYKILFKISSLVQHGCI PGPALVY 292
DB 293 QEERVRWPPRIRNEPCRGEPVSDHFFCIHHKEGIS--FEIMFLVNSVLRGVFNQFOLTE 350
QY 293 YFFRLV--DPRRNVAIEHALEKLYIKCCYD-PYRWLTEQYDGYLK-----GRQ 341
DB 351 RFFDLNRNQPKDVIASLXHL-----CTYKRPV-----FDYKRLKLVQEWTKN 395
QY 342 PPKSPSITLDGLVVRVRLVTPCKVFCGPEVNVSNRVNLANYSEDIDNLFVSVFVDEW 401
DB 396 KLLGSHQSESDISEIRLVITPRAVCLPPELVSNRVLRRYKAVAEFLRVTFMDESM 455
QY 402 EKLYSTDL-----LPKASTGSGV--RTNIYERILSLRKGFGVIGDKKFFELAFSSSLR 453
DB 456 QTINSNVLSYFVAPIVKDLTSSFSQKTYVEKRVKSYLTDGFKLCGRKYSFLAFSAQLR 515
QY 454 DNSVMKASRPGLTANDIRAMWGDFSQIKNVAKYARALGQSGSRSTLSVLRHEIV-I 512
DB 516 DRSAWFFAEDEKTRVSDIKTWGFEKD-KNVAKCAARMGLCFSTYATVDVWMPHEVTEV 574
QY 513 PDVKVHGTSYVFSIGDKIGSDFAHRAVASKGL--QVTPSAFQIRYGGYKGVGVGVDPSS 570
DB 575 PDIERNG--YVFSDDIGITITPDLAGVMEKLDVHVSFPCAYQIRYAGFGVAVRWFSKS 632
```

```
QY 141 YKQLSYENIWQVLRHPYQONAFLLQLFGAPRIYKRLNSCYSFKE-----TPDDQ 195
D 275 YKLHIFGVIKELCFMVGCEHRQALVITLRDPPVAYKRDVS--KTFGEDRLTWSNDL 332
QY 196 WYRTTDFPSSGLSSQLQRRGVRPNPESEFFHVAERENITLTGTFFV--SQK 253
D 333 WRVVDISP--GLDVS-----KNPVS-----AENHQYIDGRWLTWYIELDQ 374
QY 254 SALV-----PNVQP----- 262
D 375 STRVMDVQVQYLLDNWLRKLTFTFPEPLPNQPKVWDFLDYRQVQVSSRSWNSDES 434
QY 263 ---PBGISIPYKILFKISSLVOHCICPGPALNVYFFLVDPRRNVACIEHALEK--LY 316
D 435 LLAAPRISLPDVRQLEVCISQGIN-----EHNDRDFLE 472
QY 317 YKECC-----YDFRWLTQYDGYLKGRO--PPK-----SPSITLDDGLV 355
D 473 KLMEFCNDNSFGKDRARLILEYVADYAGKRIEFPDMELFKDHAALAYFPTFSEIPNHCA 532
QY 356 YVRVILVTPCKVYFCGPEVNVSNRVLRYNSIEDINFLRVSVFDEWE-KLYSTDL-LPKA 413
D 533 WVRVTITPRTYFTPCVPEVNRQWKAQDYFIHQFTDEVLEGRKSGEALP-- 590
QY 414 STGSGVRTNIYRILSTLRKGEVIGDKKFEFLAFSSQLRDNVSMVFASRPGLTANDIRA 473
D 591 -----LFLRVYRVLEKVGMPWHWKFALFGNSQIREAGAFMCEQSNLTGDMRA 641
QY 474 WMGDFSQKNVAKYARLQSGFSSRETLVLRHEIEVDPVKYHGTSYVFSGIGKISG 533
D 642 WMGRSHIKVIAKAYARLQCFESTRIVPGIPAPRIVTIPVEKDG--FCTDGVGKISP 699
QY 534 DPAHRVASKGLOQYTPSAFYNGYKGVGVDPD--SSMKLSLRKSMKSYESDNKLDVL 592
D 700 LLAQIVAHWDSDPPPSAYQFMGCKGVLTWPDKVMEVHRKSKQEKVFAEFNGLEV 759
QY 593 GSKQYQPCVNLRLTLTSLTGVKDEVLEQKQKAVDQDAILHDSLKAQEALE-LMSPG 651
D 760 RGSQSTATLNQIITAVLSSLCVDPQVYDMEOQLSDFNAMEQKAKTALITFIDEN 819
QY 652 ENTNIKAMLCNGYKPDPAEPFLSMMLQTPFRASKLLDLTRSRIFIPNGRTMMGLDESRT 711
D 820 HMTPIAEMLAYFGMSQEPFVRTLLQLRWSIKTLKEARLNVEKSAFVLGCVDEGT 879
QY 712 LEYGQVYQFTGAGHGEFSDLLHPNNSR-----STNSNF-TLKGNNVYAKNCL 760
D 880 LK-----GHWKVEDKQVSEKLPQIFLQIPDDVNGYRVITGTVCVGRNPSL 928
QY 761 HPGDIRLVKANVRLHMKVPCVFPQKGRPHNECSGSLDGDYIFVQWQDMIP-PR 819
D 929 HPGDIRVVEADVAPALRHLRDVWVPLTGDVDPVSMGSGDLGDDDFVITWDPILLIPKER 988
QY 820 QVQPMEXPPAPSIQLDHDTVEEVEEYPTNVIVNDSLGIITANAHVVFADREPMANSDPC 879
D 989 SHPPMISEPIACKELATPTVNNLTTFVFLVMKYNPLIAHAHALADAEVGVKSPKC 1048
QY 880 KKLAEFSTAVDFPKTGVPAEIPQSLRPEKYPDFMDXPKDTYSYISERVIGKLFKRYKDKA 939
D 1049 LELASLSMAVDYVKTGVAAEFPRRLDPKTPWFHEWK-NRHTYHSVTALCKLYDMVKRET 1107
QY 940 -----PQASSTATFTROVARSYDADMEVDGFEDYIDEAFDYKTEYDNKLGNDY 990
D 1108 FDKENYQLPDRNLKTKKRALR-----DG---TLAKARRIKSQDYDTAMRVWMCQ 1156
QY 991 YGKTEAELGGIMKASKTDFDRKADAPASVAVRALRKEARAFKRRNDIDMLPKASA 1050
D 1157 LEIATEFEVWAFVMSKPRVSGDYKLDQNVRESSALKQHPKQCKKEAG--GOLLSPVSA 1215
QY 1051 WYHVYTHPTCYCYNQGLKRAH-----FISPPVCVYDQLIQIKK 1089
D 1216 MVRVYIEVRITALEF--AKQPHVRPDGRLGTRK.TPRTMPLVSPFLWFDKUGELAR 1270
```

RESULT 10

Q8TG4

```
ID Q8TG4 PRELIMINARY; PRT: 1484 AA.
AC Q8TG4;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
DE Putative RNA-dependent RNA polymerase RDP-1.
GN RDP-1.
OS Diaporthe perijuncta.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Diaporthales; Valsaceae; Diaportha.
OX NCBI_taxid=186170;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CM3407;
RT "Putative fungal RNA-dependent RNA polymerase.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF468822; AAL78034.1; -.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1484 AA; 167213 MW; 57E0AFB8D12547BD CRC64;
```

Query Match

17.6%; Score 1031; DB 3; Length 1484;
Best Local Similarity 27.9%; Pred. No. 4.5e-65;
Matches 300; Conservative 196; Mismatches 410; Indels 168; Gaps 29;

```
QY 118 EYSIQFGIGLKKEFFFLSSGSDYKQLQSYENIWQVLRHPYQONAFLLQLF-GAPRI 176
D 300 QLEQFTVNLGRY-----GCKAFKSDIFSHPKTIWGR--RENGAWALAMLLPSPKL 351
QY 177 YKRL---ENSCYSFFKETPDDQWVRTTDFP-----PSWIGLSSSLCQFRRGVRLP 224
D 352 YGRLPVSHDDQETLQWEN--DLWARTNISSEMHLRDLPAKLEKQTKWKSIEFGRTWT-- 407
QY 225 NEESFFHYAERENITLTGTFFVSKSALYVNPQPE-----GISIPKIL----- 273
D 408 -----YFLMESSISEWESVRTIADYNVKILQVDFHDLVNPGLSNAMKLLNDSRMT 460
QY 274 -----FKISLVQHGCIQPGALNVYFF-RLV----DPRRRNVACIEHA 311
D 461 LELLGSTOSYHLDFPVRYQLEVCISHLGSLSEYVNAQFLEKLISFETDRARMLEGAEA 520
QY 312 LEKLYIKCCQDVPVRYWLTQYDGYLKGROPPKSPSITLDDGLVYVVRVLTQCKVYFCG 371
D 521 NKQFPEMEMDDP-RIL--HY--WPNARIPYA-----TLVRRAVITPTTIIFKT 566
QY 372 PEVNVSNRVLRYNSIEDINFLRVSEVDE-EWEKLYSTDLPLPKASTGSGVRTNIYERILST 430
D 567 PCVELTNRLRKYSDLNDRFLRVQFTDVTGKIFS-----SQDSNKDDNLYTRVHRV 619
QY 431 LRKGVIGDKKFEFLAFSSQLRDNVSMVFASRPGLTANDIRAMWGFQIKNVAKYAR 490
D 620 MONGVITIGDRHYKFLAFNSQFRENGAFFTCETHVTCASIRNMWGFRIHRSVGKFAAR 679
QY 491 LQSGFSSRETLVLRHEIEVDPVKYHGTSYV--FSDGIGKISGDFPAHRVASKGLOYT 548
D 680 MQCQFTTRTQNGISIPKIRQIGDIRRSGDNINWNTDGVGKISVFFARMIASERDLPET 739
QY 549 PSAFOIRYGGYKGVGVDPD--SSMKLSLRKSMKSYESDNKLDVLSKQYQPCVLANRLI 607
D 740 PSCFOMRGGCKGVVWVWPDPIPANEVHIRPSQEFKAYNGLEILIKTSTAHAHLNKQVI 799
QY 608 TLLSTLGVKDEVLEQKQKAVDQDAILHDSLKAQEALELMSPGENTNLIKAMLCNGYK 667
D 800 PVLIAGVDDAVFVRLMDDELKEYDEALADSMKAGELLRSQVDENQTTLTMAEVDTFMD 859
QY 668 DAEPLSMWLTFRASKLLDLTRSRIFIPNGRTMMGLDESRTLEYGVQVFTGAGHG 727
D 860 SEEPFLWTLRLKWKVLRKLRKQKFAISVNSAMIFGVVDGTGVL-----GHS 908
QY 728 EFSDDLHPFN-----SRSTNSNFLKGNVYVYAKNCLHPGDIVRLKA 770
```



```
RC STRAIN-CMW 5588;
RA Preisig O., Wingfield M.J.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF443073; AAL38011.1;
KW RNA-directed RNA polymerase.
FT NON_TER 1
SQ SEQUENCE 1122 AA; 126794 MW; 27D672BA58E0CBE CRC64;

Query Match 16.8%; Score 987; DB 3; Length 1122;
Best Local Similarity 30.3%; Pred. No. 4.2e-62;
Matches 267; Conservative 160; Mismatches 349; Indels 104; Gaps 18;

QY 268 IPYKILELQSSVQHGCTPGPALNYYFRVLV-----DPRRNVAQIEHAKELKLYIKBCC 322
DB 92 LDFQRYLEVCISIGLDEVSNAQFLDKLNSFDADRARMLLEGVAEASIQLEHPMKIF 151

QY 323 YDPVRLTEQYDGLKQPPSPSITLDGLVYVVRVLVTPCKYVFCGPEVNVNVR 382
DB 152 DDP-----KVLHYWPNARVPVA-----TLVRRVAVITPTTIYFKTPCVELTNRLR 197

QY 383 NYSEIDNPLRVSYFDE-EWEKLYSTDLLPKASTGSGVRYTYERILSTLRKGFVIGDK 441
DB 198 KYSDLNDRFLRVQFTDETFGRKFS-----SQDSNKDDNLYIKVRVMONGISIGDRH 250

QY 442 FFLAFSSQLRDNSVMFAFPGTANDIRAWMGDFSQIKXNVAKYAAARLQSGSSRET 501
DB 251 YRFALFNSQFRNCAFFCETDHTVCDISREWMGDFRHSVKGFAARMGQCFTTTTQV 310

QY 502 LSVLRHEIEVIDV--KVHGTSYVSDGIGKISGDFAHRVASKCGLQVTPSAFYRYGY 559
DB 311 NGISIPNIRKIDDIERTQDHTWNETDGVGKISNFFAKLIASERDLPETPCFQMRIGG 370

QY 560 KGVVGVDPP--SMKLSLRSKMSKYSNLIKLDVGLWSYQPCYLNRLITLLSTLGKDE 618
DB 371 KGVLVVWDPVPSVHVVRPSQEKFAVNVLEIKTFAFSHATFNKQVPIVVALGVDSN 430

QY 619 VLEQKQKAVDQDAILHDSLAQBALEMLSPGENTNKLKMLNGCYKPDAPPEPLSMQLQ 678
DB 431 VFSVMDLDELREYDALADSMKAGELLRSQVDENQTTLTMAEMWDTFMDKSKEPFLWTLR 490

QY 679 TFRASKLLDLRTSRIFIPNGRTMGCLDESRTL-----EYGVVQVFOFT 722
DB 491 LKWCITLKLKHKFAISVSKSAMVGVVDEIGVLGSHQATGRCYNSIESLPQIFLOVP 550

QY 723 GAGHGFEDDLHPFNNSRTSNFILKGNVNAKNPCLHPGDIRVLKAVNVRAHLMVDC 782
DB 551 IEG-----SDGRSTTNYEITGICVVGVRNPSLHPGDRVVEADVPELRHLKNV 599

QY 783 VYFPQKGRPHNECSGSLDGDYFVQWDQWIPRQVQPMVEYPP---APSIOLDH-- 836
DB 600 VYFPKTDGRDIPSMCSGGMDGDDIFVYWDERLIYTE-----WDHPPLNHEDAGSSTLDKPA 655

QY 837 DVTIEVEEYFTNYIYVNSLGIANAHHVVFADREPDAMSDPCKKLAEFLSIAVDFPKTG 896
DB 656 DVTIEDVTRFFAQYMKNDLSLGIATAHFAQAQDLGSGVKNPKCIELAKLHSMVDYIKSG 715

QY 897 VPAEISQLRPKEYPDMKPKTSIVISERVIGKLFKRVKDKAPQASSTATP--TRDVARR 955
DB 716 RPAVKRHLQPRKWPHEKEKSNRSYSALGKIYDRIKIEEFHAAVEMPPDARILSKY 775

QY 956 SYDADMEVDGFEDYIDEAFDYKTEYDNLGNLMDYIGIK--TEAEILSGGIMKASTFDR 1014
DB 776 QLEADT-----LAKASKIRATYDIAMRRLMGQHEAPVTEFEIWSFTILSRPVGSDY 827

QY 1015 KDAEALSVAVRALRKARAWFR-----RNDIDMLPKAS-----AWYHVTYHP 1058
DB 828 KLOENYGRLEAALKLRFECMEBAVTGNTQTEGFAPASSVNVLEKLDREFAAMVTTVTHND 887

QY 1059 TYWGCYNQGLKRAH-----FISPPWCYDIOI 1087
DB 888 VRAALRERSMPKPNGBGGGSEIQMPLISFPWLFHRELARY 927
```

RESULT 13

```
Q14227 PRELIMINARY; PRT; 1215 AA.
AC Q14227;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBLrel. 06, Last annotation update)
DE Hypothetical 139.5 kDa protein C6P12.09 in chromosome 1.
GN SPAC6P12.09.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Badcock K.; Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO C.ELEGANS F10B5.7, F26A3.3 AND
CC P.CHRYSOPORIUM HYPOTHETICAL 109.5 KDA PROTEIN (Q01869).
EMBL: Z98533; CAB11093.1;
KW Hypothetical protein.
SQ SEQUENCE 1215 AA; 139476 MW; 389B95C8217CB05C CRC64;

Query Match 16.2%; Score 948; DB 3; Length 1215;
Best Local Similarity 31.3%; Pred. No. 3.2e-59;
Matches 243; Conservative 155; Mismatches 316; Indels 62; Gaps 18;

QY 355 VYVRVLVTPCKYVFCGPEVNVNRLNRYSEDIDNPLRVSYFVDEWEKLYSTDLLPKAS 414
DB 452 IHKKLVTPPTLTDIVEDSLSEAGNRVIRNKDFANRFRVQITDEYKQ-----KIR 503

QY 415 TGS-GVRT-NIYERILSTLRKGFVIGDKFEFFLAFSSQLRDNSVMFAFSPGLTANDIR 472
DB 504 GSGDGRNEKLYSRIQQLLYIGIKVGNQIYEFVLAFGNSQLREHGAYFFASGDLNAKQIR 563

QY 473 AWMGDFQIKNVAKYAAARLQSGFSSRETLSVLRHEIEVIPDVKVHGTSYVFSDDGIGKIS 532
DB 564 EWMGDFSEINVSXYAARMGQCFTTKE---INRFVDISLQDDIVRNHCFITDGVGMAS 620

QY 533 GDFAHRAVASKC-GIQYTPSAFYQIRYGGYKGVGVGVDPPDSSMK-----LSLRKMSKYESD 585
DB 621 LSVIRRLSLEKVNHDMPFSAFQFRMGYKGVLSLAPPTKLEYHOGNVLPFRSQDKPKSF 680

QY 586 NIKLDVLGWSYQPCYLNRLITLLSTLGVKDEVLQKQKAVDQDAILHDSLAQBALEAL 645
DB 681 RSTLEVIKISRFSNAHLNMQLITLLEGIGVEKTVFLELTRSQLSKMNEINSKOKSILML 740

QY 646 -ELMSPGENTNKLKMLNCGYKPDAPPEFLSMMLQTFRASKILLDLRTSRIFIPNGRTMMG 704
DB 741 RDNVDEYHSTIIADFIQAGFLERDDAFTENLLNLYEVLRLIKEKQKYSVPKGAYLIG 800

QY 705 CLDESRTL-----YGVVQVQFTGAGHGEFSDDLHPFNNSRTN--SNFTLKGNV 753
DB 801 VADETGTLKGHYDDAVLSVPEIFIQT-----DTSTSFSGYSTGKLTTRVIVGLCI 851

QY 754 VAKNCPCLHPGDIRVLKAVNVRAHLMVDCVYVFPQKGRPHNECSGSLDGDYFVQWDQ 813
DB 852 VARPSLHPGDRVCKAVRCDELMLKLVIVPTTGDRSIPAMCSGGDLGDDEYTVWDQ 911

QY 814 DMIPPRQVQP--MBYPPAPSTQ-LDHVDVTIEVEEYFTNYIVNDSLGIANAHHVFA--- 867
DB 912 RLLPKIVNYPPLLESSPKKSIDFLEGKPLDSVKEFFVNYIKYDLSGLISNAKAWAHDH 971

QY 868 DREPDAMSDPCKKLAEFLSIAVDFPKTPGPAEIPSQLRPKEYPDMKPKTSIVISERV 927
DB 972 DNNPEGIFGNVLELAEMHRSKAVDFAKSGVACKMQAKYHPKYPDFMQKTKTRFSRSETA 1031

QY 928 IGLFRKRVKDKAPQASSTATFTRVARRSYDADMEVDGFE-DYIDEAFDYKTEVDNLGN 986
DB 1032 VGKIFRYAARQRESGREATY-NPIMNTVYDPCMKLPREFKTEYLNVAEVEKKHYDNDLRS 1090
```



```
QY 987 LMDYGIKTEAELSGIM---KASTFDRKDAEIAISVAVRALKEARAWKRRNDIDD 1043
Db 1091 IMARFDISTEYVTAFILEKDLAKTVNVEYGLREEVSQFDLLKKKYQYELERKALSN 1150
QY 1044 MLPKASAWYHVTYH---PTWYCYNGQLKRAH-----FISPPWCYVDQLIQIKK 1089
Db 1151 QSAFDSSEYBERINSAVAATYDVTYDQVKSVGVNGITEVLISFPYLFSSRLCOLSR 1206

RESULT 14
Q19285
ID Q19285 PRELIMINARY; PRT; 1780 AA.
AC Q19285; Q22232;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE F10B5.7 protein.
GN F10B5.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Burton J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sims M.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z66500; CAA91312.1; .
DR EMBL; Z49334; CAA91312.1; JOINED.
DR EMBL; Z49334; CAA88315.1; .
DR EMBL; Z66500; CAA88315.1; JOINED.
SQ SEQUENCE 1780 AA; 203145 MW; 16444F29CE902C2 CRC64;

Query Match 15.9%; Score 932; DB 5; Length 1780;
Best local similarity 25.1%; Pred. No. 8.4e-58;
Matches 322; Conservative 199; Mismatches 465; Indels 296; Gaps 44;

QY 14 LSAEVVSKFLEKYTG-YGTVCALVQKSGKGSRAFAKV--QFAD-NISADKIIFLANRL 69
Db 270 LFSEFASFTSRITGMLHQVLEPKM---HTLFTITPOHMDINISATAGNCPSNGL 326
QY 70 YFGSSYLKAWEMKTDIVOLRAYVQMDGITLNFQCIQSDDKFAVLGSTE----- 118
Db 327 FLVRGDFISQENTVCSVKLQSH-----HNADASRENSFKVAGSNKYLSYARFEHDK 378
QY 119 --VSIQFGLKKTFFLSSG--SADYKLOLSYENIQVHLRYPYQNAQFLLQLFCAP 174
Db 379 RLAVYVFGVRLAE---FADGDHAGFRNLNLYNLFVRIVVDMSH-ETTNSIYIQMNPP 434
QY 175 RIYKRL-ENSCYSFFK-----ETPDQWVRTTDFP-----PSWIGLS 210
Db 435 HLMGIPKNTIFHPSKSKVLNMT-CIEMTRVLSWPGDAEGRGVGCTSEAFSSQSWIRLT 493
QY 211 SSLCLQFRGRVRLNFEESFYHAERENNTIQTGFTFFVSKSALVPNPVPPGISTPY 270
Db 494 MR-----KDDNDVSSTQLMDIVTRLSA-----RSKA 521
QY 271 KILFKISSILVOHCIPGPAL-----NYFFRLVDPRRNRVAC----- 307
Db 522 KVMFGSIFSIRKKLAPSPAFHSLGSRANYALQALITRGSVFTQLDFATDENIPSSDND 581
QY 308 -----IEHALEKLYY-----IKEC-----CYDPVRWLTB-- 331
Db 582 NDEDDDDVDDTKKQWELVHEPLFLKLVRRGMKESQATEETLEQLNNAFDERRQIDVYT 641
QY 332 -----QYDYLKGRQPPKSPSITLDDGL-----VYVRRVLVTPCKVYFCGP 372
Db 642 APTTWYQSRKIOYERLLGES-----LDQVGLAKPLPKNCVSAKVIPTPSRILLMAP 694
QY 373 EVNVSNRVLRNYSIDNLFVRSFVDEEWKLYSTDLLPKASTGSGVRTNIYERILSTLR 432
```

```
Db 695 EYVAVNRVVRREGPDY--ALRCVFRDDNLGLAIRDF--SINNIDHMSNIVTEGIIYLTK 750
QY 433 KGFVIGDKKFEELAFSSQLDRNSVWMAFARPEGL-----TANDIRAMGDFSIKQVA 485
Db 751 NGIOVADRYISFLGWSNSQMRQGCYLXAPRYNALTGEVTGVEDIRVWVGDFRAISVP 810
QY 486 KYAARLGQSGFSSRETLSVLRLHEIPIVDKVH-----GTSYVFSDDGIG 529
Db 811 KMSRMGQCFTQOPTVYSSVKNIHVENIQVRLRHHWIVPEDIEGGVKNYCYSDGCG 870
QY 530 KISGDFAHRVASKGLOYPSPAFQIRYGGYKGVWGVDPD-----SSMKLSLRKSKSY-- 582
Db 871 RISIKLATHISKILOLKEVPACFQVRKFGKGLIVDPTIDDIINMPKVIFRKSQKFG 930
QY 583 ---ESDNIKDLVLSKYPQCVNLNQLITLLTLGLGVKDEVLEQK-----OKEAYD 629
Db 931 GGGELODEYLEVVKYAMPSPVCLNRPFIILDOVSEKQSSASHRRITNRVHVYLERELCS 990
QY 630 QLDAILHDSLKAOELALMSPGENTNL-----KAMLCNGYKPDABFFLSMLQTRASKL 685
Db 991 LSNMLINENQAEEV-----NRTNLAIDWNAASKRAGFELSVDPIRLDMFSIYRYNI 1044
QY 686 LDLRTRSRIFIPN--CRTHMGCLDESRTLEYGVQVQFTGAGHGEFSDDLHPFNNSRSTN 743
Db 1045 IHHISKAKIFLPPSLGRSMYGVVDETLQYGVFIQYS-----PSIRQTS 1090
QY 744 SNEFLK-GNVVAKNPCLHPGDIRVLKAVNRALHMDVCFVFPQKGRPHNECSGSDL 802
Db 1091 NRPILKTGKVLITKNPCHVPGDVRFDAVWQPALHLVDVWVFPQHPRPHDEMAGSDL 1150
QY 803 DGIYFVQVQDMIPRQVOPMEYPPAPSIQLDHDTVIEVEYFTNYIVNDSLGIIANA 862
Db 1151 DGEYSIIMQEMLLDYNEEAMVFPSSAAEEDKETTDDMVEFFLRYLQDQSIGRMSHA 1210
QY 863 HVYFADREPDMANSDPCKLAELFSTAVDFPKTGPAPETPSQLRPKEY-PDFMDKPKDKTS 921
Db 1211 HLAYADLHG--LFHENCHALAKCAVAVDFPKSGVPAELSSFEQCEMTPDYMMSSGKPM 1268
QY 922 YISERVIGLFRKVKQKAPQASSIATFTRDVARRSYD-----ADMEV-DGFEDYIDEAFD 975
Db 1269 YSTRNLNGOLHRRAR-KVEEVLEEFETRGVSFEYEDYKLCIPEDVDVFFGNEIKLVOTLT 1327
QY 976 YKTEYNKLGNDYXGKIETAEILSGIMKASKTFDRKDK-----AEAISVAVRALR 1028
Db 1328 LRDEYVDMQQLDEYGEIDEASVSGHRAISIKRLAGMERDDYSFYHTDKVVELRYEKL 1387
QY 1029 KEARAWF-----KRRNDIDMLPKASAWYHVTY-HPTY---WGCYNQ 1067
Db 1388 AVFRAKFFEEFGCEINIENDGKNTRLKCTKAMHEKIRQWYFVAYVQPKINKAGRCIGQ- 1446
QY 1068 LKRAHPTISFPWCYVDQLIQIKK 1089
Db 1447 -----SLFWAVMDALCULRR 1461

RESULT 15
Q9BH56
ID Q9BH56 PRELIMINARY; PRT; 1579 AA.
AC Q9BH56;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE M01G12.12 protein.
GN M01G12.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McIay K.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
```



```

[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2016(1998).
[3]
RN SEQUENCE FROM N.A.
RA Lennard N.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132848; CAC35915.1;
DR EMBL: Z81571; CAC35915.1; JOINED.
DR EMBL: Z81571; CAC35914.1;
DR EMBL: AL132848; CAC35914.1; JOINED.
SQ SEQUENCE 1579 AA; 182870 MW; B64517200715E26C CRC64;

Query Match 15.3%; Score 898; DB 5; Length 1579;
Best Local Similarity 25.1%; Pred. No. 2e-55;
Matches 317; Conservative 196; Mismatches 449; Indels 302; Gaps 45;

QY 43 GSRAFAKVOFADNISADKIITLANNLYFGS-----SYLKAWEKTDIVOLRA----- 90
Db 101 GSDFWAK-----TLDVIEVPLAAMFYGNIGQYTIYISHMOISFSGKKISADANELLNK 153
QY 91 -----YVDQMGITLNFQCGISDDKEAVLGSSTEVSIQFGIGLKKFFFLSSGSADYKQL 145
Db 154 IVAEFEDRADMTIVFQC-----LKRKRINQITII 184
QY 146 SYENINOVLUHRYGQNAQFLLQLFGAPRI-----YKRLN-SCYSFF 188
Db 185 RKTIRIIVDPOVMNKTIRHFLACPLPIROGSDVDDDKPSTQKPFYKRTNRYSCIG-T 243
QY 189 KETPDDQWRTTDFPPSWIGLSSSLC-----LOPRRGVRL--PNFEE- 228
Db 244 KEYGSPHESAISDPTTIELQKESDNGSDNDLYRLVLSRLSRTGVQIEFANFPKV 303
QY 229 -----SFFHZAERNITLTGTTFFVFSQKSAVLPVNPQPEG:SIYPKILFKLSSLV 280
Db 304 DVPIGKVPYLYRPTSSKSAFECFYNCPPSKMKIIDAQISINENDGROFAITYLIECLL 363
QY 281 QHGCIPI-----GPALNV---YFRLVDPERRNVACIEHALEKLYIYKE----- 320
Db 364 SRGAIVKQVLTDBICWQFLGLITHYL-----ENDKLC EALEDLIYLDGKRIG 416
QY 321 ---OCYDVP--RWLTEQYDGLKGRQPPKSPSITLDDGLVYVRRVLVTPCKYVFCGPEVN 375
Db 417 SIWKCFHKIQCKRLVMQLTNGMSEQE-----IEEGYQVRKVIPTPRVIYTPPEMI 468
QY 376 VSNRVLRYNSEDIDNFRVSVFDEWEKLYSTDLILPKASTGSGVRTNIYERILST----- 430
Db 469 MGNRVLRNFRDGTGTHVLRVTRDDNNRKM-----RANATGELLIDICVKKY 513
QY 431 LRKGFVIGDKKFEPLAFSSQLRDNVSMFPAS-----RPLGTANDIRAW 474
Db 514 LEHGIVVANRDFGLGCGSSQMRDNGAYFVWKNTDNRHKNACKMNSKFKENI--DSVRNQ 571
QY 475 MGDFSQTKNKAAYARLGQSGFSRET-LSV-----LRHEIIVDPKVHGTSYVFSOG 527
Db 572 LGNFIQIENIPKLMARLGQCFTQSRLTGVSIGLDPDNYCLTHDLS--GGRSSNGSEYTFSDG 629
QY 528 IGKISGDFAHFVASKGL-OYTPSAFOIRYGGYGVGVGDP-----DSS 570
Db 630 VGMMSYEFQAQEVSLMFGSRVSPCFQFRYRGMKGVLAIIDPLDKERNWQEKHGISLSKN 689
QY 571 MKLSLRKMSKSYSDNIKLDVLGNSKYOP---CYLNPQLITLSTLGKDEVLEQKQKEA 627
Db 690 IKCVFRPSQIKFEKGQILGDQVEMVKYSSPVLVALNKLINIL-----DQVSEMOSLEC 743
QY 628 VQDLDAIHLSDLSKAE---ALELSPGENTNILKAMLN-----CGYKPDABPFLS 674
Db 744 HRRITSRIEELMDLOTLSFAKQMTDEACFRNKLKEFFPRIDIVLRTINGFTLSNEPFR 803
```

```

QY 675 MMLQT---FRASKLLDLRSTRIPN--GFTMMGCLDESRTLEYGOVFOFTGAGHGEF 729
Db 804 SLVKASIKFAITKQL---FKEQIQIPSELGRSMGLGVVDETGILQYGVFOYT-KNHRNI 859
QY 730 SDDLHPFNNSRSTNSNFIKGNVVVAKNCPCLHPGDIRVLKAVNVRALHMHVDCVVFQPKG 789
Db 860 ---LPPRDSNRKVLGSEIVTGTVLLTKNPGIVPGDVRIFEAVDIPELHHLCDVVVFPQHG 916
QY 790 KRPHNECSGDLDDGDIYFVCWDQDMIPPROVQPMVEYPPAPSIQLDHDVTLEE-VEEYET 848
Db 917 PRPHDEMAGSDLDGDEYSVINDQKLLERNEEAFDFAVENKLOTYEWEDIDDLMRDVIYV 976
QY 849 NVIVNDSLGIITANAHVVPADREPDMAUSDCKKLAELFSAVDFPKTGVPAAE----- 900
Db 977 EYLKDLVGLIANSHLHNSDOYG--LTSRVCMNLAKSKQAVDFSKSGKPPDELQTTWKT 1034
QY 901 -----IPSQIRPKEYPDF-MDKPDKTSYISERVIGKLFKRVKDKAPQASSIATFRDV 952
Db 1035 DDATGEMIPPE-RAERVPDYHVGSDDHMPKYVSPRLCGKLFREFOG-IDNAMKISSEKSEQ 1092
QY 953 ARRSYDADMEVDGPFEDYIDEAFDYKTEYDNKLGMLNDYIGIKTEAEILSGGIMK-ASKTF 1011
Db 1093 YKIEVDESIRVDGFEEYEMEDAKKOLASYNQKLSMTMDTYGIOSEGEIMSGCIIEMNRIS 1152
QY 1012 DRKDD-----AEAISVAVRALRKEARAW-----FKRR 1038
Db 1153 DSDQDDMSFYNTNRMETKMTALVSKFTIFFQOFGSQEVCTLLPDAYNESNFFNCFE 1212
QY 1039 NDIDDLMPKASAWYHVTHYTYWGCYN--OGLKRAHFISFPWCYVDQLIQIKKDKARNRP 1096
Db 1213 NENEIRKKAWAYR-----ACYECAKSTREPRKLSFAWIAYDVTAIKETKVLNV 1264
QY 1097 VLNL 1100
Db 1265 EMNI 1268
```

Search completed: November 6, 2002, 03:46:07
Job time : 49 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:37:53 ; Search time 167 Seconds
(without alignments)
92.517 Million cell updates/sec

Title: US-09-782-874-2
Perfect score: 1114
Sequence: 1 MGKTIQVGFYLLSAEVVK.....RPVLNLSRLAQLSHRLVLK 1114

Scoring table:
OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 90412 seqs, 13869272 residues

Word size : 0
Total number of hits satisfying chosen parameters: 90412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1114	100.0	1114	10	US-09-782-874-2 Sequence 2, Appl
2	218	19.6	218	10	US-09-782-874-3 Sequence 3, Appl
3	17	1.5	17	10	US-09-782-874-12 Sequence 12, Appl
4	14	1.3	14	10	US-09-782-874-10 Sequence 10, Appl
5	14	1.3	14	10	US-09-782-874-11 Sequence 11, Appl
6	10	0.9	13	10	US-09-782-874-13 Sequence 13, Appl
7	8	0.7	283	10	US-09-764-846-220 Sequence 220, App
8	8	0.7	477	10	US-09-764-846-140 Sequence 140, App
9	8	0.7	506	10	US-09-801-368-214 Sequence 214, App
10	7	0.6	18	10	US-09-864-761-45788 Sequence 45788, A
11	7	0.6	24	10	US-09-864-761-46295 Sequence 46295, A
12	7	0.6	46	10	US-09-764-870-336 Sequence 336, App
13	7	0.6	46	10	US-09-764-869-1017 Sequence 1017, App
14	7	0.6	51	10	US-09-925-299-1317 Sequence 1317, App
15	7	0.6	85	10	US-09-864-761-37705 Sequence 37705, A
16	7	0.6	157	10	US-09-741-569-427 Sequence 427, App
17	7	0.6	249	10	US-09-205-658-309 Sequence 309, App
18	7	0.6	279	10	US-09-789-582A-2 Sequence 2, Appl
19	7	0.6	322	10	US-09-747-835A-54 Sequence 54, Appl

333 10 US-09-815-242-11987 Sequence 11987, A
358 10 US-09-801-368-148 Sequence 148, App
403 10 US-09-870-379-2 Sequence 2, Appl
436 8 US-08-980-068B-19 Sequence 19, Appl
469 10 US-09-925-301-1377 Sequence 1377, App
482 10 US-09-815-242-13863 Sequence 13863, A
536 10 US-09-815-242-4978 Sequence 4978, App
563 10 US-09-815-242-10841 Sequence 10841, A
575 10 US-09-815-242-10240 Sequence 10240, A
612 8 US-08-910-386A-2 Sequence 2, Appl
737 10 US-09-771-161A-195 Sequence 195, App
834 10 US-09-934-909-29 Sequence 29, Appl
1025 8 US-08-910-386A-7 Sequence 7, Appl
8 10 US-09-872-349-12 Sequence 12, Appl
10 10 US-09-124-280A-19 Sequence 19, Appl
14 10 US-09-765-527-129 Sequence 129, App
14 10 US-09-881-490-90 Sequence 90, Appl
14 10 US-09-881-490-101 Sequence 101, App
20 10 US-09-864-761-43391 Sequence 43391, A
23 10 US-09-860-793-14 Sequence 14, Appl
25 10 US-09-864-761-38599 Sequence 38599, A
26 10 US-09-860-793-13 Sequence 13, Appl
27 10 US-09-864-761-45443 Sequence 45443, A
32 10 US-09-864-761-41081 Sequence 41081, A
37 10 US-09-864-761-47440 Sequence 47440, A
38 10 US-09-764-869-810 Sequence 810, App
39 10 US-09-864-761-45869 Sequence 45869, A
39 10 US-09-864-761-46284 Sequence 46284, A
44 10 US-09-864-761-36412 Sequence 36412, A
45 10 US-09-864-761-40514 Sequence 40514, A
45 10 US-09-864-761-45158 Sequence 45158, A

ALIGNMENTS

RESULT 1
US-09-782-874-2
; Sequence 2, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassenegeger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
POLYPEPTIDES HAVING THE ENZYMACTIC ACTIVITY OF AN
RNA-DIRECTED RNA POLYMERASE (RDRP)

NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/782,874
FILING DATE: 08-Feb-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/811,583
FILING DATE: 05-MAR-1997

ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1

TELECOMMUNICATION INFORMATION:

Db 1 RTMGCLESTLEYGVQVFTGAGHGEFSDDLHPFNNSRSTNSFLKGNVYVAKNPC 60
QY 760 LHPGDIRVLKAVNRALHHMVDVVFQKGRPHPNCEGSGDLGGDIYFVCWDDQDMIPPR 819
Db 61 LHPGDIRVLKAVNRALHHMVDVVFQKGRPHPNCEGSGDLGGDIYFVCWDDQDMIPPR 120
QY 820 QVQPMPEYPPASIQLDHVDVTEEEYFTNTIVNDSIGIIANAHVVFADREPDMAUSDPC 879
Db 121 QVQPMPEYPPASIQLDHVDVTEEEYFTNTIVNDSIGIIANAHVVFADREPDMAUSDPC 180
QY 880 KKLAELEFIADVDFPKTGVPAEIPSOIRPKPEYDFMDKP 917
Db 181 KKLAELEFIADVDFPKTGVPAEIPSOIRPKPEYDFMDKP 218
RESULT 3
US-09-782-874-12
; Sequence 12, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassenegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-782-874-12

Query Match 1.5%; Score 17; DB 10; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.6e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 331 EQYDGYLKGROPPKSPS 347
Db 1 EQYDGYLKGROPPKSPS 17

RESULT 4
US-09-782-874-10
; Sequence 10, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassenegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-782-874-10
Query Match 1.3%; Score 14; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.7e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 377 SNRYLRNYSIEDIDN 390
Db 1 SNRYLRNYSIEDIDN 14
RESULT 5
US-09-782-874-11
; Sequence 11, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassenegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:

```
;
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-782-874-11
;
; Query Match 1.3%; Score 14; DB 10; Length 14;
; Best Local Similarity 100.0%; Pred. No. 5.7e-07;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1007 ASKTFDRRKDAEAI 1020
; |||||
; DB 1 ASKTFDRRKDAEAI 14
;
; RESULT 6
; US-09-782-874-13
; Sequence 13, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassensgger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
;
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDRP)
;
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
;
;
;
; ADDRESSSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-782-874-13
;
; Query Match 0.9%; Score 10; DB 10; Length 13;
; Best Local Similarity 100.0%; Pred. No. 0.0045;
; Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 784 VFQKQKGRPH 793
; |||||
; DB 1 VFQKQKGRPH 10
;
; RESULT 7
; US-09-764-846-220
; Sequence 220, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 220
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (158)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (252)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-846-220
;
; Query Match 0.7%; Score 8; DB 10; Length 283;
; Best Local Similarity 100.0%; Pred. No. 6;
; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 739 SRSTNSNF 746
; |||||
; DB 114 SRSTNSNF 121
;
; RESULT 8
; US-09-764-846-140
; Sequence 140, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```

```

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 140
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-846-140

```

```

Query Match      0.7%; Score 8; DB 10; Length 477;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 739 SRSTNSNF 746
Db 114 SRSTNSNF 121

```

```

RESULT 9
US-09-801-368-214
; Sequence 214, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 214
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-214

```

```

Query Match      0.7%; Score 8; DB 10; Length 506;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 923 ISERVIGK 930
Db 314 ISERVIGK 321

```

```

RESULT 10
US-09-864-761-45788
; Sequence 45788, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.

```

```

; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45788
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC022317.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.67
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.76
US-09-864-761-45788

```

```

Query Match      0.6%; Score 7; DB 10; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 572 KLSLRKS 576
Db 12 KLSLRKS 18

```

```

RESULT 11
US-09-864-761-46295
; Sequence 46295, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.

```

```
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46295
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138761.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
US-09-864-761-46295
```

```
Query Match 0.6%; Score 7; DB 10; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 209 LSSSLCL 215
Db 7 LSSSLCL 13
```

RESULT 12

```
US-09-764-870-336
; Sequence 336, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
```

```
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 336
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-870-336
```

```
Query Match 0.6%; Score 7; DB 10; Length 46;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 207 IGLSSSL 213
Db 28 IGLSSSL 34
```

RESULT 13

```
US-09-764-869-1017
; Sequence 1017, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1017
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-869-1017
```

```
Query Match 0.6%; Score 7; DB 10; Length 46;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 569 SSMKLSL 575
Db 30 SSMKLSL 36
```

RESULT 14

```
US-09-925-299-1317
; Sequence 1317, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1317
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
```

FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (14)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (16)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (28)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (35)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (38)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1317

Query Match 0.6%; Score 7; DB 10; Length 51;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 KKFFFL 134
| | | | |
DB 41 KKFFFL 47

RESULT 15
US-09-864-761-37705
; Sequence 37705, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864.761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37705
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF067844.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: O08586, EVALUE 2.00e-44
; OTHER INFORMATION: EST_HUMAN HIT: A1222037.1, EVALUE 2.00e-43
US-09-864-761-37705

Query Match 0.6%; Score 7; DB 10; Length 85;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 LKAQAL 645
| | | | |
DB 67 LKAQAL 73

RESULT 16
US-09-741-669-427
; Sequence 427, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA-009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-427

Query Match 0.6%; Score 7; DB 10; Length 157;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11987
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11987

Query Match 0.6%; Score 7; DB 10; Length 333;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 832 IQLDHDV 838
|||||
Db 115 IQLDHDV 121

RESULT 21
US-09-801-368-148
; Sequence 148, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,557
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 148
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-148

Query Match 0.6%; Score 7; DB 10; Length 358;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 QLITLIS 611
|||||
Db 347 QLITLIS 353

RESULT 22
US-09-870-379-2
; Sequence 2, Application US/09870379
; Patent No. US20020150954A1

; GENERAL INFORMATION:
; APPLICANT: Donald L. Durden
; APPLICANT: ADVANCED RESEARCH & TECHNOLOGY INSTITUTE
; TITLE OF INVENTION: Compositions and Methods for Identifying
; FILE REFERENCE: ARTI 0024-US
; CURRENT APPLICATION NUMBER: US/09/870,379
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: PCT/US01/17358
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/274/167
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/208,437
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-870-379-2

Query Match 0.6%; Score 7; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAEAL 645
|||||
Db 146 LKAEAL 152

RESULT 23
US-08-980-068B-19
; Sequence 19, Application US/08980068B
; Patent No. US20020081718A1
; GENERAL INFORMATION:
; APPLICANT: HOTTA, Yoshiaki
; TITLE OF INVENTION: A POLYPEPTIDE COMMON TO GLIAL CELLS MISSING (GCM)
; FILE REFERENCE: 97-1513*/LC/00653
; CURRENT APPLICATION NUMBER: US/08/980,068B
; CURRENT FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-980-068B-19

Query Match 0.6%; Score 7; DB 8; Length 436;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 898 PAEIPSQ 904
|||||
Db 193 PAEIPSQ 199

RESULT 24
US-09-925-301-1377
; Sequence 1377, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 1694
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1377
 ; LENGTH: 469
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-925-301-1377

Query Match 0.6%; Score 7; DB 10; Length 469;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 SFFKETP 192
 |||||
 Db 225 SFFKETP 231

RESULT 25

US-09-815-242-13863
 ; Sequence 13863, Application US/09815242
 ; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13863
 ; LENGTH: 482
 ; TYPE: PRT
 ; ORGANISM: Salmonella typhi

; NAME/KEY: VARIANT
 ; LOCATION: (1)...(482)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-815-242-13863

Query Match

Best Local Similarity 0.6%; Score 7; DB 10; Length 482;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030
 |||||
 Db 399 VRLRKE 405

RESULT 26

US-09-815-242-4978
 ; Sequence 4978, Application US/09815242

; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4978
 ; LENGTH: 536
 ; TYPE: PRT
 ; ORGANISM: Enterococcus faecalis
 US-09-815-242-4978

Query Match 0.6%; Score 7; DB 10; Length 536;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ERAWFK 1036
 |||||
 Db 183 ERAWFK 189

RESULT 27

US-09-815-242-10841
 ; Sequence 10841, Application US/09815242
 ; Patent No. US20020061569A1

; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10841
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10841

Query Match 0.68; Score 7; DB 10; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ERAWFK 1036
|||||||
Db 210 ERAWFK 216

RESULT 28
US-09-815-242-10240
; Sequence 10240, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/181,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10240
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10240

Query Match 0.68; Score 7; DB 10; Length 575;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030
|||||||
Db 399 VRLRKE 405

RESULT 29
US-08-910-386A-2

; Sequence 2, Application US/08910386A
; Patent No. US2002092041A1
; GENERAL INFORMATION:
; APPLICANT: Ronald, Pamela C.
; APPLICANT: Wang, Guo-Liang
; APPLICANT: Song, Wen-Yuang
; APPLICANT: Hulbert, Scott
; APPLICANT: Richter, Todd
; TITLE OF INVENTION: Procedures and Materials for Conferring
; TITLE OF INVENTION: Disease Resistance in Plants
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,386A
; FILING DATE: 13-AUG-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-05895005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 612 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-910-386A-2

Query Match 0.68; Score 7; DB 8; Length 612;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1099 NLSSLRA 1105
|||||||
Db 246 NLSSLRA 252

RESULT 30
US-09-771-161A-195
; Sequence 195, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 195
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-771-161A-195

Query Match 0.6%; Score 7; DB 10; Length 737;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 647 LMSPGN 653
Db 386 LMSPGN 392

RESULT 31

US-09-934-909-29
Sequence 29, Application US/09934909
Patent No. US20020058275A1
GENERAL INFORMATION:
APPLICANT: FISHEL, Richard A.
APPLICANT: GRADIA, Scott
APPLICANT: ACHARYA, Samir
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR EFFECTING ADENINE
TITLE OF INVENTION: NUCLEOTIDE MODULATION OF DNA MISMATCH RECOGNITION
TITLE OF INVENTION: PROTEINS
FILE REFERENCE: 9855-6U3
CURRENT APPLICATION NUMBER: US/09/934,909
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: US 60/093,935
PRIOR FILING DATE: 1998-07-23
PRIOR APPLICATION NUMBER: US 60/066,977
PRIOR FILING DATE: 1997-11-28
PRIOR APPLICATION NUMBER: US 60/057,136
PRIOR FILING DATE: 1997-08-28
PRIOR APPLICATION NUMBER: US 09/143,571
PRIOR FILING DATE: 1998-08-28
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 29
LENGTH: 834
TYPE: PRT
ORGANISM: Homo sapiens
US-09-934-909-29

Query Match 0.6%; Score 7; DB 10; Length 834;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 728 EFSDDLH 734
Db 366 EFSDDLH 372

RESULT 32

US-08-910-386A-7
Sequence 7, Application US/08910386A
Patent No. US20020092041A1
GENERAL INFORMATION:
APPLICANT: Ronald, Pamela C.
APPLICANT: Wang, Guo-Liang
APPLICANT: Song, Wen-Yuang
APPLICANT: Hulbert, Scott
APPLICANT: Richter, Todd
TITLE OF INVENTION: Procedures and Materials for Conferring
TITLE OF INVENTION: Disease Resistance in Plants
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,386A
FILING DATE: 13-AUG-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-05869500S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1025 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-910-386A-7

Query Match 0.6%; Score 7; DB 8; Length 1025;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1099 NLSSLRA 1105
Db 246 NLSSLRA 252

RESULT 33

US-09-872-349-12
Sequence 12, Application US/09872349
Patent No. US20020132980A1
GENERAL INFORMATION:
APPLICANT: Sidney Pestka
TITLE OF INVENTION: PHOSPHORYLATED PROTEINS AND USES RELATED THERETO
FILE REFERENCE: PBLI-P01-007
CURRENT APPLICATION NUMBER: US/09/872,349
CURRENT FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 60/208,240
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 60/255,296
PRIOR FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 12
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: phosphorylated
OTHER INFORMATION: peptide
US-09-872-349-12

Query Match 0.5%; Score 6; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 7.5e+04;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1100 LSSLRA 1105
Db 3 LSSLRA 8

RESULT 34

US-09-124-280A-19
Sequence 19, Application US/09124280A
Patent No. US20020034520A1
GENERAL INFORMATION:
APPLICANT: Porro, Massimo
TITLE OF INVENTION: VACCINES FOR PREVENTION OF GRAM-
TITLE OF INVENTION: NEGATIVE BACTERIAL INFECTIONS AND ENDOTOXIN RELATED DISEASE
NUMBER OF SEQUENCES: 45

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,280A
FILING DATE: July 29,1998
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-124-280A-19

Query Match      0.5%; Score 6; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 KRFEFF 133
DB 4 KRFEFF 9

RESULT 35
US-09-765-527-129
Sequence 129, Application US/09765527
Patent No. US2002006638A1
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
TITLE OF INVENTION: Methods for Recombinant Microbial Production of
Fusion Proteins and BPI-Derived Peptides
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,527
FILING DATE: 18-Jan-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/621,803
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27129/33199

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,280A
FILING DATE: July 29,1998
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-124-280A-19

Query Match      0.5%; Score 6; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 KRFEFF 133
DB 4 KRFEFF 9

RESULT 35
US-09-765-527-129
Sequence 129, Application US/09765527
Patent No. US2002006638A1
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
TITLE OF INVENTION: Methods for Recombinant Microbial Production of
Fusion Proteins and BPI-Derived Peptides
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,527
FILING DATE: 18-Jan-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/621,803
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27129/33199

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,280A
FILING DATE: July 29,1998
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 129:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: "XMP.250"
FEATURE:
NAME/KEY: Modified-site
LOCATION: C-Terminus
OTHER INFORMATION: /label= Amidation
/Note- "The C-Terminus is Amidated."
SEQUENCE DESCRIPTION: SEQ ID NO: 129:
US-09-765-527-129

Query Match      0.5%; Score 6; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 LLIQLF 171
DB 6 LLIQLF 11

RESULT 36
US-09-881-490-90
Sequence 90, Application US/09881490
Patent No. US20020077298A1
GENERAL INFORMATION:
APPLICANT: Little II, Roger G.
Lim, Edward
Fadem, Mitchell B.
TITLE OF INVENTION: Anti-Fungal Peptides
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th FloorDrive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/881,490
FILING DATE: 14-Jun-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/119,858
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/372,105
FILING DATE: 13-JAN-95
APPLICATION NUMBER: 08/306,473
FILING DATE: 15-SEP-94
APPLICATION NUMBER: 08/273,540
FILING DATE: 11-JUL-94
APPLICATION NUMBER: 08/209,762
FILING DATE: 11-MAR-94
APPLICATION NUMBER: 08/193,222
FILING DATE: 14-JAN-94
APPLICATION NUMBER: 08/093,202
FILING DATE: 15-JUL-93
APPLICATION NUMBER: 08/030,644
FILING DATE: 12-MAR-93
ATTORNEY/AGENT INFORMATION:
```

NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 100-238/11021US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/707-8889
TELEFAX: 312/707-9155
TELEX: 650 388-1248
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: "XMP.250"
FEATURE:
NAME/KEY: Modified-site
LOCATION: C-Terminus
OTHER INFORMATION: /label= Amidation
/note= "The C-Terminus is Amidated"
SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-09-881-490-90
Query Match 0.5%; Score 6; DB 10; Length 14;
Best Local Similarity 100.0%; Pred.No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 166 LLIQLF 171
Db 6 LLIQLF 11
RESULT 37
US-09-881-490-101
Sequence 101, Application US/09881490
Patent No. US20020077298A1
GENERAL INFORMATION:
APPLICANT: Little II, Roger G.
Lim, Edward
Fadem, Mitchell B.
TITLE OF INVENTION: Anti-Fungal Peptides
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th Floor Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/881,490
FILING DATE: 14-Jun-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/119,858
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/372,105
FILING DATE: 13-JAN-95
APPLICATION NUMBER: 08/306,473
FILING DATE: 15-SEP-94
APPLICATION NUMBER: 08/273,540
FILING DATE: 11-JUL-94
APPLICATION NUMBER: 08/209,762
FILING DATE: 11-MAR-94
APPLICATION NUMBER: 08/183,222
FILING DATE: 14-JAN-94
APPLICATION NUMBER: 08/093,202
FILING DATE: 15-JUL-93

APPLICATION NUMBER: 08/030,644
FILING DATE: 12-MAR-93
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 100-238/11021US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/707-8889
TELEFAX: 312/707-9155
TELEX: 650 388-1248
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: "XMP.261"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /label= N-methyl-L
/note= "Position 6 is N-Methyl-leucine"
FEATURE:
NAME/KEY: Modified-site
LOCATION: C-Terminus
OTHER INFORMATION: /label= Amidation
/note= "The C-Terminus is Amidated"
SEQUENCE DESCRIPTION: SEQ ID NO: 101:
US-09-881-490-101
Query Match 0.5%; Score 6; DB 10; Length 14;
Best Local Similarity 100.0%; Pred.No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 166 LLIQLF 171
Db 6 LLIQLF 11
RESULT 38
US-09-864-761-43391
Sequence 4331, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 43391
LENGTH: 20
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006385.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.85
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.83
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.77
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.84
US-09-864-761-43391
Query Match 0.5%; Score 6; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 902 PSQLRP 907
DB 6 PSQLRP 11
RESULT 39
US-09-860-793-14
Sequence 14, Application US/09860793
Patent No. US20020136734A1
GENERAL INFORMATION:
APPLICANT: Pruett, John H
APPLICANT: Temeyer, Kevin B
APPLICANT: Kunz, Sidney E
APPLICANT: Fisher, William F
TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic
FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.
CURRENT APPLICATION NUMBER: US/09/860.793
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/366,603
PRIOR FILING DATE: 1999-08-03
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 14
LENGTH: 23
TYPE: PRT
ORGANISM: Psoroptes ovis
US-09-860-793-14

Query Match 0.5%; Score 6; DB 10; Length 23;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 YTGyGT 31
|||||

DB 9 YTGyGT 14

RESULT 40

US-09-864-761-38599
Sequence 38599, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 38599
LENGTH: 25
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 283844.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
OTHER INFORMATION: SWISSPROT HIT: P55194, EVALUATE 2.00e-01
US-09-864-761-38599

Query Match 0.5%; Score 6; DB 10; Length 25;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DPDSM 571
|||||


```
Db 20 DPDSM 25

RESULT 41
US-09-860-793-13
; Sequence 13, Application US/09860793
; Patent No. US20020136734A1
; GENERAL INFORMATION:
; APPLICANT: Pruett, John H
; APPLICANT: Temeyer, Kevin B
; APPLICANT: Kunz, Sidney E
; APPLICANT: Fisher, William F
; TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic
; FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.
; CURRENT APPLICATION NUMBER: US/09/860,793
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/366,603
; PRIOR FILING DATE: 1999-08-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 13
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Psoroptes ovis
US-09-860-793-13

Query Match 0.5%; Score 6; DB 10; Length 26;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 VTGYGT 31
|||||
Db 21 YIGYGT 26

RESULT 42
US-09-864-761-45443
; Sequence 45443, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663

RESULT 43
US-09-864-761-41081
; Sequence 41081, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
```

; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41081
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC021677.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; US-09-864-761-41081

Query Match 0.5%; Score 6; DB 10; Length 32;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FGPPYL 13
|||||
Db 18 FGPPYL 23

RESULT 44
US-09-864-761-47440
; Sequence 47440, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47440
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL132857.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.51
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.44
; OTHER INFORMATION: SWISSPROT HIT: P77891, EVALUE 4.90e+00
; US-09-864-761-47440

Query Match 0.5%; Score 6; DB 10; Length 37;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 RVLRYN 384
|||||
Db 21 RVLRYN 26

RESULT 45
US-09-764-869-810
; Sequence 810, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 810
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-869-810

Query Match 0.5%; Score 6; DB 10; Length 38;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 ETLSVL 505
|||||
Db 26 ETLSVL 31

RESULT 46
US-09-864-761-45869

; Sequence 45869, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Aeonica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 45869

; LENGTH: 39

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC016948.4

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93

; OTHER INFORMATION: EST_HUMAN HIT: BE880522.1, EVALUE 7.20e+00

US-09-864-761-45869

Query Match 0.5%; Score 6; DB 10; Length 39;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 502 LSVLRH 507

Db 19 LSVLRH 24

|||||

RESULT 47

US-09-864-761-46284

; Sequence 46284, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

; FILE REFERENCE: Aeonica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 46284

; LENGTH: 39

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC008045.5

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4

; OTHER INFORMATION: EXPRESSED IN HUMAN HIT: BE348399.1, EVALUE 9.00e-07

; OTHER INFORMATION: SWISSPROT HIT: P55200, EVALUE 4.30e-01

US-09-864-761-46284

Query Match 0.5%; Score 6; DB 10; Length 39;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 688 LRTRSR 693

Db 19 LRTRSR 24

|||||

Db 2 LRTRSR 7

US-09-864-761-36412

Query Match 0.5%; Score 6; DB 10; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 688 LRTRSR 693
Db 7 LRTRSR 12

RESULT 48

US-09-864-761-36412

; Sequence 36412, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36412
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008045.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: SWISSPROT HIT: P55200, EVALUE 3.10e-01
; OTHER INFORMATION: EST_HUMAN HIT: AW135326.1, EVALUE 9.00e-09

US-09-864-761-40514

; Sequence 40514, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40514
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL139819.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1

OTHER INFORMATION: MAP TO ALL37853.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.56
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.73
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
OTHER INFORMATION: EST_HUMAN HIT: BE733841.1, EVALUATE 4.00e+00
US-09-864-761-45158

Query Match 0.5%; Score 6; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 902 PSQLRP 907
Db 17 PSQLRP 22

Search completed: November 6, 2002, 03:43:47
Job time : 175 secs

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
OTHER INFORMATION: EST_HUMAN HIT: BF061077.1, EVALUATE 1.00e-10
US-09-864-761-40514

Query Match 0.5%; Score 6; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LSAEVV 19
Db 29 LSAEVV 34

RESULT 50

US-09-864-761-45158
Sequence 45158, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Reomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 45158
LENGTH: 45
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:35:08 ; Search time 81 Seconds

(without alignments)

1322.145 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 1114

Sequence: 1 MGKTIQVGFPPYLLSAEVK.....RPVLNLSLRAQLSHRLVLK 1114

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 50 summaries

Database : PIR_73:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1114	100.0	1114	2 T30819	RNA-directed RNA p
2	42	3.8	1116	2 T30828	RNA-directed RNA p
3	14	1.3	1133	2 T01920	probable RNA-direc
4	9	0.8	583	2 H59165	hypothetical prote
5	8	0.7	230	2 T22763	hypothetical prote
6	8	0.7	265	2 S60947	hypothetical prote
7	8	0.7	286	2 A95395	protein [imported
8	8	0.7	341	2 T38874	probable short cha
9	8	0.7	352	2 H37272	histidinol-phospha
10	8	0.7	386	2 S31436	probable membrane
11	8	0.7	506	2 S69045	protein kinase MKK
12	8	0.7	641	2 G28771	hypothetical prote
13	8	0.7	830	2 T04848	protein kinase hom
14	8	0.7	901	2 D70116	transcription elon
15	8	0.7	1101	2 G70951	probable ATP-depen
16	8	0.7	1385	2 S42330	156k protein - Pla
17	8	0.7	1564	2 T27121	hypothetical prote
18	7	0.6	29	2 AB0717	hypothetical prote
19	7	0.6	48	2 A57125	polyphenolic adhes
20	7	0.6	66	2 AB2371	hypothetical prote
21	7	0.6	70	2 S68954	polyphenolic adhes
22	7	0.6	79	2 A01030	probable membrane
23	7	0.6	89	2 T46384	hypothetical prote
24	7	0.6	91	2 T07231	hypothetical prote
25	7	0.6	112	2 S14355	glutathione transf
26	7	0.6	116	2 AG2329	50S ribosomal prot
27	7	0.6	121	2 D97840	hypothetical prote
28	7	0.6	122	2 AC2172	hypothetical prote
29	7	0.6	124	2 AG1754	bacteriophage prot

30 7 0.6 141 2 H71504
31 7 0.6 142 2 A81663
32 7 0.6 145 2 H82567
33 7 0.6 149 2 B89329
34 7 0.6 151 2 A60943
35 7 0.6 154 2 S75019
36 7 0.6 157 2 H64930
37 7 0.6 160 2 F83541
38 7 0.6 167 2 T14262
39 7 0.6 169 2 D84722
40 7 0.6 183 1 QRECEB
41 7 0.6 183 2 A98931
42 7 0.6 183 2 E85779
43 7 0.6 183 2 S04743
44 7 0.6 197 2 T15923
45 7 0.6 206 2 A84865
46 7 0.6 206 2 E75310
47 7 0.6 207 2 F83815
48 7 0.6 211 1 JC2368
49 7 0.6 211 2 S23753
50 7 0.6 213 2 T22984

ribosomal protein
ribosomal protein
outer membrane pro
hypothetical prote
lg heavy chain pre
glutathione peroxi
hypothetical prote
probable glutathio
glutathione peroxi
probable glutathio
vitamin B12 transp
vitamin B12 transp
vitamin B12 transp
TPA-induced protei
hypothetical prote
probable glutathio
hypothetical prote
hypothetical prote
ribosomal protein
ribosomal protein
hypothetical prote

ALIGNMENTS

RESULT 1

T30819

RNA-directed RNA polymerase (EC 2.7.7.48) - tomato

C:Species: Lycopersicon esculentum (tomato)

C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000

C:Accession: T30819

R:Schluebel, W.; Fellissier, T.; Riedel, L.; Thalmair, S.; Schliebel, R.; Kempe, D.; Lo

Plant Cell 10, 2087-2102, 1998

A:Title: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato.

A:Reference number: Z20885; MUID:99055198; PMID:9836747

A:Accession: T30819

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1114 <SCH>

A:Cross-references: EMBL:Y10403; NID:g4038591; PIDN:CAA71421.1; PID:g4038592

C:Genetics:

A:Note: RdRP

C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase

C:Keywords: nucleotidyltransferase

Query Match 100.0%; Score 1114; DB 2; Length 1114;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGKTIQVGFPPYLLSAEVKSFLEKTYGTVCALEVKQSKGGSRAFAKVFADNISADK 60

Db 1 MGKTIQVGFPPYLLSAEVKSFLEKTYGTVCALEVKQSKGGSRAFAKVFADNISADK 60

Qy 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDMGDGITLNFQGISDDKFAVLGSTEVS 120

Db 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDMGDGITLNFQGISDDKFAVLGSTEVS 120

Qy 121 IQFGIGLKPPFFLLSSGSADYKQLSYENIWVYVHRPYGNAQFLLIQLFGAPRIYKRL 180

Db 121 IQFGIGLKPPFFLLSSGSADYKQLSYENIWVYVHRPYGNAQFLLIQLFGAPRIYKRL 180

Qy 181 ENSCYSFKEKTPDDQWVTTDFPPSWIGLSSSLCQFRGVRVLPNFESFFHYAERENNI 240

Db 181 ENSCYSFKEKTPDDQWVTTDFPPSWIGLSSSLCQFRGVRVLPNFESFFHYAERENNI 240

Qy 241 TLQGTGFFVFSQKALVPNVOPPEGISIPYKILFKISLVHGGICPGALNVYFRLVDP 300

Db 241 TLQGTGFFVFSQKALVPNVOPPEGISIPYKILFKISLVHGGICPGALNVYFRLVDP 300

Qy 301 RRRNVACIEHALEKLYIKKCYDVPVRLTEQYDGYLKGROPPKSPSITLDDGLVYVRRV 360

Db 301 RRRNVACIEHALEKLYIKKCYDVPVRLTEQYDGYLKGROPPKSPSITLDDGLVYVRRV 360

QY 361 LVTCKVYFCGPEVNVNRLVNYSEDINFLRVSVFDEWEKLYSTDLLPKASTGSGVR 420
Db 361 LVTCKVYFCGPEVNVNRLVNYSEDINFLRVSVFDEWEKLYSTDLLPKASTGSGVR 420
QY 421 TNYIERILSTLRKGVIGDKFEFLAFSSQLRDNVWVFASRGLTANDIRAWMGDFSQ 480
Db 421 TNYIERILSTLRKGVIGDKFEFLAFSSQLRDNVWVFASRGLTANDIRAWMGDFSQ 480
QY 481 INWAKYARLQSGFSGSRETLVSRHEIEVDPVKVHGTSTYVFSDDGKISGDFAHRYA 540
Db 481 INWAKYARLQSGFSGSRETLVSRHEIEVDPVKVHGTSTYVFSDDGKISGDFAHRYA 540
QY 541 SKCGLYTPSAFQIRYGGYGVVDPDSSMKLSLRKSMKSYESDNILKDLVGLMSKYOPC 600
Db 541 SKCGLYTPSAFQIRYGGYGVVDPDSSMKLSLRKSMKSYESDNILKDLVGLMSKYOPC 600
QY 601 YLNROLITLLSTGLVKDEVLEQKQEAVDQDLAILHDSLKQAQEALELMSPGENTINILKAM 660
Db 601 YLNROLITLLSTGLVKDEVLEQKQEAVDQDLAILHDSLKQAQEALELMSPGENTINILKAM 660
QY 661 LNCGYKPAEPFLSMWLOTFRASKLLDLTRSRIPFIPNGRTMNGCLDSRTLEYGQVFEVQ 720
Db 661 LNCGYKPAEPFLSMWLOTFRASKLLDLTRSRIPFIPNGRTMNGCLDSRTLEYGQVFEVQ 720
QY 721 FTGAGHGFEFDDLHPFNNSRSTNSFILKGNVYVAKNCPCLHPGDIRVLKAVNVRALHVMV 780
Db 721 FTGAGHGFEFDDLHPFNNSRSTNSFILKGNVYVAKNCPCLHPGDIRVLKAVNVRALHVMV 780
QY 781 DCWVFPQKGRPHNECSGSDLDGDIYFCWDQDMIPPRQVQPMYPPAPSIQLDHDVTI 840
Db 781 DCWVFPQKGRPHNECSGSDLDGDIYFCWDQDMIPPRQVQPMYPPAPSIQLDHDVTI 840
QY 841 EEVEEYFTNYIVNDSIGLIANAHHVVFADREPDMASDCKKLAELFSAVDFFPKTGPAP 900
Db 841 EEVEEYFTNYIVNDSIGLIANAHHVVFADREPDMASDCKKLAELFSAVDFFPKTGPAP 900
QY 901 IPSQLRPEYDFMDKPKDTSYISERVIGKLFKRVKDKAPOASSIATATRDVARRSDAD 960
Db 901 IPSQLRPEYDFMDKPKDTSYISERVIGKLFKRVKDKAPOASSIATATRDVARRSDAD 960
QY 961 MEVDGFEDYIDAFDYKTEYDNKGLNMDYGIKTEAEILSGGIMKASKTFDRKDAEAI 1020
Db 961 MEVDGFEDYIDAFDYKTEYDNKGLNMDYGIKTEAEILSGGIMKASKTFDRKDAEAI 1020
QY 1021 SVAVRALKEARAWFKRRNDIDMLPKASAWYHVTYHPTYWGVCYNQGLKRAHFFSFPCV 1080
Db 1021 SVAVRALKEARAWFKRRNDIDMLPKASAWYHVTYHPTYWGVCYNQGLKRAHFFSFPCV 1080
QY 1081 YDQLIQIKKDKARNRPVNLSSRLAQLSHRLVVK 1114
Db 1081 YDQLIQIKKDKARNRPVNLSSRLAQLSHRLVVK 1114

RESULT 2
T30828
RNA-directed RNA polymerase - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000
C:Accession: T30828
R:Schiebel, W.; Pelissier, T.; Riedel, L.; Thalmeir, S.; Schiebel, R.; Kempe, D.; Lottsp
submitted to the EMBL Data Library, October 1998
A:Description: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from Tom
A:Reference number: Z20890
A:Accession: T30828
A:Molecule type: mRNA
A:Residues: 1-1116 <SCH>
A:Cross-references: EMBL:AJ011576; PIDN:CAA09697.1
C:Genetics:
A:Note: RdRP
C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase

Query Match 3.8%; Score 42; DB 2; Length 1116;
Best Local Similarity 100.0%; Pred. No. 4.3e-36;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 979 EYDNKGLNMDYGIKTEAEILSGGIMKASKTFDRKDAEAI 1020
Db 981 EYDNKGLNMDYGIKTEAEILSGGIMKASKTFDRKDAEAI 1022

RESULT 3
T01920
probable RNA-directed RNA polymerase (EC 2.7.7.48) - Arabidopsis thaliana
N:Alternate names: protein T22B4.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 10-Dec-1999
C:Accession: T01920; T08192
R:Strong, C.; Graves, T.; Duckels, G.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of A. thaliana F2P3.
A:Reference number: Z14455
A:Accession: T01920
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1133 <STR>
A:Cross-references: EMBL:AF080120; NID:g3600045; PID:g3600048
A:Experimental source: cultivar Columbia
R:Bevan, M.; Pohl, T.; Weizengger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Sch
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16098
A:Accession: T08192
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1133 <BEV>
A:Cross-references: EMBL:AL049876
A:Experimental source: cultivar Columbia; BAC clone T22B4
C:Genetics:
A:Map position: 4
A:Introns: 184/3; 820/2; 870/3
A:Note: F2P3.11; T22B4.110
C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase
C:Keywords: nucleotidyltransferase

Query Match 1.3%; Score 14; DB 2; Length 1133;
Best Local Similarity 100.0%; Pred. No. 9.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 549 PSAFQIRYGGYKGV 562
Db 578 PSAFQIRYGGYKGV 591

RESULT 4
H69165
hypothetical protein MTH500 - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 04-Mar-2000
C:Accession: H69165
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jilwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: H69165
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-583 <MTH>
A:Cross-references: GB:AE000833; GB:AE000666; NID:g2621559; PIDN:AAB85006.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH500
C:Superfamily: Methanobacterium thermoautotrophicum hypothetical protein MTH500

Query Match 0.8%; Score 9; DB 2; Length 583;
Best Local Similarity 100.0%; Pred. No. 1.4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 345 SPSITLDDG 353
|||||
DB 221 SPSITLDDG 229

RESULT 5

hypothetical protein F56A8.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22763

R:McMurray, A.

submitted to the EMBL Data Library, December 1996

A:Reference number: Z19612

A:Accession: T22763

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-230 <WTL>

A:Cross-references: EMBL:Z83230; PIDN:Z83230; PIDN:CAB05742.1; GSPDB:GN00021; CESP:F56A8.2

A:Experimental source: clone F56A8

C:Genetics:

A:Gene: CESP:F56A8.2

A:Map position: 3

A:Introns: 65/2; 97/3

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 230;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 DSLKAQEA 644

|||||

DB 156 DSLKAQEA 163

RESULT 6

S60947

hypothetical protein YOR220w - yeast (Saccharomyces cerevisiae)

A:Alternate names: hypothetical protein O5050; hypothetical protein YOR50-10

C:Species: Saccharomyces cerevisiae

C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002

C:Accession: S60947; S67113; S71722

R:Galissou, F.; Dujon, B.

submitted to the EMBL Data Library, October 1995

A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome

A:Reference number: S60938

A:Accession: S60947

A:Molecule type: DNA

A:Residues: 1-265 <GAL>

A:Cross-references: EMBL:X92441; NID:g1050762; PID:g1050772

R:Boyer, J.; Fairhead, C.; Gailion, L.; Galissou, F.; Michaux, G.; Thierry, A.; Dujon, B.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67104

A:Accession: S67113

A:Molecule type: DNA

A:Residues: 1-265 <BOY>

A:Cross-references: EMBL:Z75128; NID:g1420509; PID:g1420510; MIPS:YOR220w

A:Experimental source: strain S288C

R:Galissou, F.; Dujon, B.

Yeast 12, 877-885, 1996

A:Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV

A:Reference number: S71722

A:Accession: S71722

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-265 <GAW>

A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63183.1; PID:g1050772

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Cross-references: SGD:S0005746

A:Map position: 15B

C:Superfamily: Saccharomyces hypothetical protein YOR220w

Query Match 0.7%; Score 8; DB 2; Length 265;

Best Local Similarity 100.0%; Pred. No. 8.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 PPKSPSIT 349

|||||

DB 252 PPKSPSIT 259

RESULT 7

A95395

protein [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSymA

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: A95395

R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; E

; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli

A:Reference number: A95262; MUID:21396509; PMID:11481432

A:Accession: A95395

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-286 <KUR>

A:Cross-references: GB:AE006469; PIDN:AAK65723.1; PID:g14524217; GSPDB:GN00165

A:Experimental source: strain 1021, magaplasmid pSymA

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Sma1945

A:Genome: plasmid

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 286;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 SKLIDLRT 690

|||||

DB 251 SKLIDLRT 258

RESULT 8

T38874

probable short chain dehydrogenase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T38874

R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z21732

A:Accession: T38874

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-341 <CON>

A:Cross-references: EMBL:Z69727; PIDN:CAA93565.1; GSPDB:GN000066; SPDB:SPAC4G9.15

A:Experimental source: strain 972h-; cosmid c4G9

C:Genetics:

A:Gene: SPDB:SPAC4G9.15

A:Map position: 1

A:Introns: 5/1; 48/3

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 341;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1025 RALRKEAR 1032

Db 327 RALRKEAR 334

RESULT 9

H97272

histidinol-phosphate aminotransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: H97272

R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: H97272

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-352 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80971.1; PID:g15026090; GSPDB:GNO0168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3031

C:Superfamily: histidinol phosphate aminotransferase

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 352;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 IDNFLRVS 395

Db 326 IDNFLRVS 333

RESULT 10

S51436

probable membrane protein YLR191w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein L9470.1

C:Species: Saccharomyces cerevisiae

C:Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 06-Feb-1998

C:Accession: S51436

R:Wohldmann, P.

submitted to the EMBL Data Library, November 1994

A:Description: The sequence of S. cerevisiae cosmid 9470.

A:Reference number: S51414

A:Accession: S51436

A:Molecule type: DNA

A:Residues: 1-386 <WOH>

A:Cross-references: EMBL:U17246; NID:g577192; PID:g577193; MIPS:YLR191w

C:Genetics:

A:Gene: SGB:PEX13

A:Cross-references: SGD:S0004181; MIPS:YLR191w

A:Map position: 12R

C:Superfamily: SH3 homology

C:Keywords: transmembrane protein

F:264-280/Domain: transmembrane #status predicted <TMM>

F:313-367/Domain: SH3 homology <SH3>

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 386;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VFGFPYLL 14

Db 272 VFGFPYLL 279

RESULT 11

S69045

protein kinase MKK2 (EC 2.7.1.1) - yeast (Saccharomyces cerevisiae)

N:Alternate names: mitogen-activated protein kinase homolog; protein LPI6c; pr

C:Species: Saccharomyces cerevisiae

C:Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 24-Sep-1999

C:Accession: S69045; B48069; S30786

R:Hall, J.; DePaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.

submitted to The EMBL Data Library, December 1995

A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A:Reference number: S69040

A:Accession: S69045

A:Molecule type: DNA

A:Residues: 1-506 <HAL>

A:Cross-references: EMBL:U43703; NID:g1244769; PIDN:AA68220.1; PID:g1244775; MIPS:YP

R:Rie, K.; Takase, M.; Lee, K.S.; Levin, D.E.; Araki, H.; Matsumoto, K.; Oshima, Y.

Mol. Cell. Biol. 13, 3076-3083, 1993

A:Title: MKK1 and MKK2, which encode Saccharomyces cerevisiae mitogen-activated prote

A:Reference number: A48069; MUID:93233568; PMID:8386320

A:Accession: B48069

A:Molecule type: DNA

A:Residues: 1-251, 'S', 253-506 <IRI>

A:Cross-references: EMBL:D13785; NID:g218442; PIDN:BA02933.1; PID:dl003439; PID:g218

A:Note: sequence extracted from NCBI backbone (NCBIP:129897)

C:Genetics:

A:Gene: SGD:MKK2; SSP33

A:Cross-references: SGD:S0006061; MIPS:YPL140C

A:Map position: 16L

C:Function:

A:Description: phosphotransferase; serine/threonine-specific protein kinase; involved

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

C:Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase; signal

F:212-481/Domain: protein kinase homology <KIN>

F:220-228/Region: protein kinase ATP-binding motif

F:342/Active site: Asp #status predicted

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 506;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 923 ISERVIGK 930

Db 314 ISERVIGK 321

RESULT 12

G28771

hypothetical protein C2814 (photosynthetic gene cluster) - Rhodobacter capsulatus

C:Species: Rhodobacter capsulatus

C:Date: 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change 04-Sep-1998

C:Accession: G28771

R:Youvan, D.C.; Bylina, E.J.; Alberti, M.; Begusch, H.; Hearst, J.E.

Cell 37, 949-957, 1984

A:Title: Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-

A:Reference number: A90850; MUID:84259352; PMID:6744416

A:Accession: G28771

A:Molecule type: DNA

A:Residues: 1-641 <YOU>

A:Cross-references: GB:K01183

C:Superfamily: hypothetical protein C2814

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 641;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 GLTANDIR 472

Db 615 GLTANDIR 622

RESULT 13

T04848

protein kinase homolog F16G20.10 - Arabidopsis thaliana

N:Alternate names: protein F16G20.10; protein F21p8.200

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 17-Mar-2000

C:Accession: T04848; T05369
 R:Bevan, M.; Massenet, O.; Clabault, G.; Quigley, F.; Mache, R.; Bancroft, I.; Mewes, H.
 submitted to the Protein Sequence Database, August 1998
 A:Reference number: Z15386
 A:Accession: T04848
 A:Molecule type: DNA
 A:Residues: 1-830 <BEV>
 A:Cross-references: EMBL:AL022347
 A:Experimental source: cultivar Columbia; BAC clone F21P8
 R:Bevan, M.; Watson, M.D.; Gibbons, T.; Bartley, J.; Bancroft, I.; Mewes, H.W.; Mayer, K.
 submitted to the Protein Sequence Database, August 1998
 A:Reference number: Z15413
 A:Accession: T05369
 A:Molecule type: DNA
 A:Residues: 618-830 <BEW>
 A:Cross-references: EMBL:AL031326
 A:Experimental source: cultivar Columbia; BAC clone F16G20
 C:Genetics:
 A:Map position: 4
 A:Introns: 484/1; 524/3; 595/1; 674/2; 725/3
 A:Note: F21P8.200; F16G20.10
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 0.7%; Score 8; DB 2; Length 830;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 606 LITLSTL 613
 |||||
 Db 162 LITLSTL 169

RESULT 14
 D70116
 transcription elongation factor (greA) homolog - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 15-Oct-1999
 C:Accession: D70116
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Keriavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 A:Accession: D70116
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-901 <KLE>
 A:Cross-references: GB:AE001124; GB:AE000783; NID:g2688003; PIDN:AAC66506.1; PID:g268800
 A:Experimental source: strain B31
 C:Keywords: transcription factor

Query Match 0.7%; Score 8; DB 2; Length 901;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 764 DIRVLKAV 771
 |||||
 Db 359 DIRVLKAV 366

RESULT 15
 G70951
 Probable ATP-dependent DNA helicase - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: G70951
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, S.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gene
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: G70951
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1101 <COL>
 A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16666.1; PID:el2
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv3201c

Query Match 0.7%; Score 8; DB 2; Length 1101;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1021 SVAVRALR 1028
 |||||
 Db 374 SVAVRALR 381

RESULT 16
 S34230
 156K protein - Plantago asiatica mosaic virus
 C:Species: Plantago asiatica mosaic virus
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Nov-1999
 C:Accession: S34230
 R:Atabekov, J.G.
 submitted to the EMBL Data Library, February 1993
 A:Reference number: S34230
 A:Accession: S34230
 A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-1395 <ATA>
 A:Cross-references: EMBL:Z21647; NID:g311644; PIDN:CAA79761.1; PID:g311645
 C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

Query Match 0.7%; Score 8; DB 2; Length 1385;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 425 ERILSTLR 432
 |||||
 Db 597 ERILSTLR 604

RESULT 17
 T27121
 hypothetical protein Y53C10A.9 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T27121
 R:White, S.
 submitted to the EMBL Data Library, November 1998
 A:Reference number: Z20314
 A:Accession: T27121
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1564 <WIL>
 A:Cross-references: EMBL:AL033536; PIDN:CAA22142.1; CESP:Y53C10A.9
 A:Experimental source: clone Y53C10A
 C:Genetics:
 A:Gene: CESP:Y53C10A.9
 A:Introns: 43/3; 92/2; 149/2; 226/3; 354/1; 712/3; 817/1; 875/1; 916/3; 984/3; 1069/2
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 0.7%; Score 8; DB 2; Length 1564;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 606 LITLSTL 613
 |||||
 Db 254 LITLSTL 261

RESULT 18
 AB0717
 hypothetical protein STY1874a [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
 A:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AB0717
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Cronin, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; et al.
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AB0717
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-29 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD02107.1; PID:g16502942; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY1874a

 Query Match 0.6%; Score 7; DB 2; Length 29;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 990 YGKTE 996
 |||||
 DB 8 YGKTE 14

 RESULT 19
 A57125
 polyphenolic adhesive protein 3 - blue mussel
 C:Species: Mytilus edulis (blue mussel)
 C:Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jan-2000
 C:Accession: A57125
 R:Papov, V.V.; Diamond, T.V.; Biemann, K.; Waite, J.H.
 J. Biol. Chem. 270, 20183-20192, 1995
 A:Title: Hydroxyarginine-containing polyphenolic proteins in the adhesive plaques of the mussel Mytilus edulis
 A:Reference number: A57125; MUID:95378278; PMID:7650037
 A:Accession: A57125
 A:Molecule type: protein
 A:Residues: 1-48 <PAP>
 F:3,4,8,14,19,22,25,29,32,47/Modified site: 3',4'-dihydroxyphenylalanine (Tyr) #status experimental
 F:12,13,21,24,27,28,41,43,44/Modified site: 4-hydroxyarginine (Arg) #status experimental

 Query Match 0.6%; Score 7; DB 2; Length 48;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 555 RVGGYKG 561
 |||||
 DB 28 RVGGYKG 34

 RESULT 20
 AB2371
 hypothetical protein asr4522 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AB2371
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AB2371
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-66 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA076221.1; PID:g17133658; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: asr4522

Query Match 0.6%; Score 7; DB 2; Length 66;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 VQPPEGI 266
 |||||
 DB 55 VQPPEGI 61

RESULT 21

S68954

polyphenolic adhesive protein 3A precursor - Mediterranean mussel

N:Alternate names: foot protein 3A

C:Species: Mytilus galloprovincialis (Mediterranean mussel)

C:Date: 23-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 04-Feb-2000

C:Accession: S68954

R:Inoue, K.; Takeuchi, Y.; Miki, D.; Odo, S.; Harayama, S.; Waite, J.H.

Eur. J. Biochem. 239, 172-176, 1996

A:Title: Cloning, sequencing and sites of expression of genes for the hydroxyarginine

A:Reference number: S68954; MUID:96305382; PMID:8706704

A:Accession: S68954

A:Molecule type: mRNA

A:Residues: 1-70 <INO>

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-70/Product: polyphenolic adhesive protein 3A #status predicted <MAT>

F:27,28,32,38,43,46,50,53,69,70/Modified site: 3',4'-dihydroxyphenylalanine (Tyr) #st

F:36,37,45,48,49,62,64,67/Modified site: 4-hydroxyarginine (Arg) #status predicted

Query Match 0.6%; Score 7; DB 2; Length 70;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RVGGYKG 561

|||||

DB 49 RVGGYKG 55

RESULT 22

AD1030

probable membrane protein STY4565 [imported] - Salmonella enterica subsp. enterica se

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AD1030

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; PMID:11677608

A:Accession: AD1030

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-79 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD09341.1; PID:g16505341; GSPDB:GN00176

C:Genetics:

A:Gene: STY4565

Query Match 0.6%; Score 7; DB 2; Length 79;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 455 SVMMEAS 462

|||||

DB 73 SVMMEAS 79

RESULT 23

T46384
 hypothetical protein DKF2p434p2119.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 02-Sep-2000
 C:Accession: T46384
 R:Ottewaeldeir, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, January 2000
 A:Reference number: Z23031
 A:Accession: T46384
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-89 <AAA>
 A:Cross-references: EMBL:AL137543
 A:Experimental source: adult testis; clone DKF2p434p2119
 C:Genetics:
 A:Note: DKF2p434p2119.1
 C:Superfamily: tropomodulin

Query Match 0.6%; Score 7; DB 2; Length 89;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 ENTN1LK 658
 |||||
 Db 48 ENTN1LK 54

RESULT 24

T07231
 hypothetical protein 91 - Chlorella vulgaris chloroplast
 C:Species: chloroplast Chlorella vulgaris
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
 C:Accession: T07231
 R:Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Nakas
 Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997
 A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga Chlo
 A:Reference number: Z15985; MUID:97303241; PMID:9159184
 A:Accession: T07231
 A:Status: preliminary; translated from GB/EMBL/DDBB
 A:Molecule type: DNA
 A:Residues: 1-91 <WAK>
 A:Cross-references: EMBL:AB001684; NID:g2224352; PIDN:BAAS7878.1; PID:g2224394
 C:Genetics:
 A:Genome: chloroplast
 C:Keywords: chloroplast

Query Match 0.6%; Score 7; DB 2; Length 91;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 ASRPGLT 467
 |||||
 Db 6 ASRPGLT 12

RESULT 25

SI4355
 glutathione transferase (EC 2.5.1.18) class alpha chain 10 - rat (fragments)
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Nov-1993 #sequence_revision 08-Nov-1996 #text_change 08-Nov-1996
 C:Accession: SI4355
 R:Meyer, D.J.; Gilmore, K.S.; Coles, B.; Dalton, K.; Hulbert, P.B.; Ketterer, B.
 Biochem. J. 274, 619, 1991
 A:Title: Structural distinction of rat GSH transferase subunit 10.
 A:Reference number: SI4355; MUID:91174781; PMID:2006926
 A:Accession: SI4355
 A:Molecule type: protein
 A:Residues: 1-29;30-51;52-60;61-66;67-94;95-112 <NEY>
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 0.6%; Score 7; DB 2; Length 112;

Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
 |||||
 Db 79 KDKARNR 85

RESULT 26

AG2329
 50S ribosomal protein L17 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. Strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AG2329
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AG2329
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-116 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA875889.1; PID:gl7133325; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: rpl17
 C:Superfamily: Escherichia coli ribosomal protein L17

Query Match 0.6%; Score 7; DB 2; Length 116;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 DKIIITLA 65
 |||||
 Db 49 DKIIITLA 55

RESULT 27

D97840
 hypothetical protein RC1124 [imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
 C:Accession: D97840
 R:Ogata, H.; Audic, S.; Repesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.
 Science 293, 2093-2098, 2001
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; MUID:21442074; PMID:11557893
 A:Accession: D97840
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-121 <KUR>
 A:Cross-references: GB:AF006914; PIDN:AAL03662.1; PID:gl5620249; GSPDB:GN00173
 C:Genetics:
 A:Gene: RC1124

Query Match 0.6%; Score 7; DB 2; Length 121;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KTIQVFG 9
 |||||
 Db 33 KTIQVFG 39

RESULT 28

AC2172
 hypothetical protein all2930 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AC2172

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AC2172
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-122 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA074629.1; PID:gl17132024; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all2930

Query Match 0.6%; Score 7; DB 2; Length 122;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 VLRNYS 386

Db 107 VLRNYS 113

|||||||

RESULT 29

AG1754

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AG1734

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihl, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunz, F.; Kurapk, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of Listeria species

A:Reference number: AB107; MUID:21537279; PMID:11679669

A:Accession: AG1734

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-124 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC97807.1; PID:gl16415102; GSPDB:GN00178

A:Experimental source: strain Clip1262

C:Genetics:

A:Gene: lin2580

Query Match 0.6%; Score 7; DB 2; Length 124;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 976 YKTEYDN 982

Db 32 YKTEYDN 38

|||||||

RESULT 30

H71504

C:Species: Chlamydia trachomatis

C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 13-Aug-1999

C:Accession: H71504; I40747

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,

Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra

A:Reference number: A71570; MUID:99000809; PMID:9784136

A:Accession: H71504

A:Molecule type: DNA

A:Residues: 1-141 <ARN>

A:Cross-references: GB:AB001323; GB:AB001273; NID:g3328931; PIDN:AAC68107.1; PID:g332894

A:Experimental source: serotype D, strain UW-3/Cx

R:Gu, L.; Wenman, W.M.; Remacha, M.; Meuser, R.; Coffin, J.; Kaul, R.

J. Bacteriol. 177, 2594-2601, 1995

A:Title: Chlamydia trachomatis RNA polymerase alpha subunit: sequence and structural ana

A:Reference number: I40743; MUID:95247702; PMID:7730299

A:Accession: I40747

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 20-115,'R',117-141 <GUL>

A:Cross-references: GB:L33834; NID:g620026; PIDN:AAA74990.1; PID:g620030

C:Genetics:

A:Gene: rli7

C:Superfamily: Escherichia coli ribosomal protein L17

C:Keywords: protein biosynthesis; ribosome

Query Match 0.6%; Score 7; DB 2; Length 141;

Best Local Similarity 100.0%; Pred. No. 61;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 RVIGKLF 932

Db 98 RVIGKLF 104

|||||||

RESULT 31

AB1663

C:Species: Chlamydia muridarum, Chlamydia trachomatis MOPN

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000

C:Accession: AB1663

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hicke

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, J.; Kolonay, J.; McClarty, G.; Salzbe

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39

A:Reference number: AB1500; MUID:20150255; PMID:10684935

A:Accession: AB1663

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-142 <RET>

A:Cross-references: GB:AE002347; GB:AE002160; NID:g7190815; PIDN:AAF39596.1; PID:g719

A:Experimental source: strain N199 (MOPN)

C:Genetics:

A:Gene: TC0793

C:Superfamily: Escherichia coli ribosomal protein L17

Query Match 0.6%; Score 7; DB 2; Length 142;

Best Local Similarity 100.0%; Pred. No. 61;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 RVIGKLF 932

Db 98 RVIGKLF 104

|||||||

RESULT 32

H82567

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: H82567

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: AB2515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82567

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-145 <SIM>

A:Cross-references: GB:AE004045; GB:AE003849; NID:g9107517; PIDN:AAF95144.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carrier

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fi

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak, A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva, M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
C;Contents: annotation
C;Genetics:
A;Gene: XF2345

Query Match 0.6%; Score 7; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 625 KEAVDQL 631
| | | | |
DB 49 KEAVDQL 55

RESULT 33

B89929
hypothetical protein SA1329 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: B89929

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: B89929
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-149 <KUR>
A;Cross-references: GB:BA000018; PID:gl3701297; PIDN:BA842591.1; GSPDB:GN00149

A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1329

C;Superfamily: ferric uptake regulator

Query Match 0.6%; Score 7; DB 2; Length 149;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 934 KVDDKAP 940
| | | | |
DB 46 KVDDKAP 52

RESULT 34

A60943
Ig heavy chain precursor V region (clone HN.14) - human

C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-1999

C;Accession: A60943; A48165
R;Desai, R.; Spatz, L.; Matsuda, T.; Ilyas, A.A.; Berman, J.E.; Alt, F.W.; Kabat, E.A.; J. Neuroimmunol. 30, 245, 1990

A;Title: Molecular cloning of a human immunoglobulin heavy chain variable (V-H) region
A;Reference number: A60943; MUID:91036050; PMID:1699376
A;Accession: A60943
A;Status: not compared with conceptual translation
A;Molecule type: mRNA

A;Residues: 1-151 <DES>
J. Neuroimmunol. 26, 35-41, 1990

A;Title: Molecular cloning of a human immunoglobulin heavy chain variable (V-H) region
A;Reference number: A48165; MUID:90094677; PMID:1688442
A;Accession: A48165
A;Molecule type: mRNA

A;Residues: 1-36, 'M', '38-62, 'AR', 67-151 <DE2>
A;Note: this sequence has been corrected in reference A60943

C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F:34-117/Domain: immunoglobulin homology <IMM>

Query Match 0.6%; Score 7; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ASTGSGV 419
| | | | |
DB 122 ASTGSGV 128

RESULT 35

S75019

glutathione peroxidase (EC 1.11.1.9) - Synechocystis sp. (strain PCC 6803)
N;Alternate names: hypothetical protein slr1992
C;Species: Synechocystis sp.
A;Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75019
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-154 <KAN>
A;Cross-references: EMBL:D90910; GB:AB001339; NID:gl652956; PIDN:BAAL7881.1; PID:gl65
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: glutathione peroxidase
C;Keywords: oxidoreductase; selenocysteine
F:34/Modified site: selenocysteine #status predicted

Query Match 0.6%; Score 7; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
| | | | |
DB 30 VASKCGL 36

RESULT 36

H64930

hypothetical protein b1720 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C;Accession: H64930
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Ross, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A54720; MUID:97426617; PMID:9278503
A;Accession: H64930

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-157 <BLAT>

A;Cross-references: GB:AE000267; GB:U00096; NID:gl788011; PIDN:ARC74790.1; PID:gl7880
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: Escherichia coli hypothetical protein b1720

Query Match 0.6%; Score 7; DB 2; Length 157;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 RQLITLL 610
| | | | |
DB 37 RQLITLL 43

RESULT 37
F83541
Probable glutathione peroxidase PA0838 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83541
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83541
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <STO>
A:Cross-references: GB:AE004518; GB:AE004091; NID:9946725; PIDN:AAG04227.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0838
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 160;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
|||||||
DB 32 VASKCGL 38

RESULT 38
T14262
glutathione peroxidase (EC 1.11.1.9) - common sunflower
C:Species: Helianthus annuus (common sunflower)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Jun-2000
C:Accession: T14262
R:Roedel-Drevet, P.; Gagne, G.; Tourvielle de Labrouhe, D.; Dufaire, J.P.; Nicolas, P.;
Physiol. Plantarum 103, 385-394, 1998
A:Title: Molecular characterization, organ distribution and stress-mediated induction of
A:Reference number: Z17562
A:Accession: T14262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-167 <ROE>
A:Cross-references: EMBL:Y14429
C:Genetics:
A:Gene: GPxha-1
C:Function:
A:Description: catalyzes reduction of hydroperoxides by glutathione, thus protecting bio
C:Superfamily: glutathione peroxidase
C:Keywords: oxidoreductase

Query Match 0.6%; Score 7; DB 2; Length 167;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
|||||||
DB 37 VASKCGL 43

RESULT 39
D84722
Probable glutathione peroxidase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: D84722
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Woffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84722
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-169 <STO>
A:Cross-references: GB:AE002093; NID:94582452; PIDN:AD24836.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g31570
A:Map position: 2
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
|||||||
DB 37 VASKCGL 43

RESULT 40
ORECBE
Vitamin B12 transport periplasmic protein btuE - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 30-Jun-1988 #sequence_revision 31-Oct-1997 #text_change 01-Mar-2002
C:Accession: F64929; B24498
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F64929
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-183 <BLAT>
A:Cross-references: GB:AE000266; GB:U00096; NID:q1787997; PIDN:AAC74780.1; PID:q17880
A:Experimental source: strain K-12, substrain MG1655
R:Friedrich, M.J.; Deveau, L.C.; Kadner, R.J.
J. Bacteriol. 167, 928-934, 1986
A:Title: Nucleotide sequence of the btuCD genes involved in vitamin B12 transport in
A:Reference number: A24498; MUID:86304184; PMID:3528129
A:Accession: B24498
A:Molecule type: DNA
A:Residues: 1-183 <FR>
A:Cross-references: GB:M14031; NID:q145441; PIDN:AAA23527.1; PID:q145444
C:Comment: This protein, thought to be located in the periplasm, is not essential for
C:Comment: This sequence is homologous with that of bovine glutathione peroxidase (41
C:Genetics:
A:Gene: btuE
A:Map position: 37 min
C:Superfamily: glutathione peroxidase
C:Keywords: vitamin B12 transport
F:37/Active site: Cys #status predicted

Query Match 0.6%; Score 7; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
|||||||
DB 33 VASKCGL 39

RESULT 41
A98931
vitamin B12 transport protein ECS2417 [imported] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: A98931
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
A:Reference number: A98629; MUID:21156231; PMID:11258796
A:Accession: A98931
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-183 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA035840.1; PID:g13361884; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
|||||||
Db 33 VASKCGL 39

RESULT 42
E85779
vitamin B12 transport [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85779
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85779
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-183 <STO>
A:Cross-references: GB:AE005174; NID:g12515721; PIDN:AAG56697.1; GSPDB:GN00145; UNCP:227
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: btuE
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
|||||||
Db 33 VASKCGL 39

RESULT 43
S04743
TPA-induced protein 11 - mouse
N:Alternate names: tetradecanoyl phorbol acetate-induced protein 11
C:Species: Mus musculus (house mouse)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 19-Apr-2002
C:Accession: S04743
R:Varum, B.C.; Lim, R.W.; Sukhatme, V.P.; Herschman, H.R.
Oncogene 4, 119-120, 1989
A:Title: Nucleotide sequence of a cDNA encoding TIS11, a message induced in Swiss 3T3 ce
A:Reference number: S04743; MUID:89128189; PMID:2915901
A:Accession: S04743
A:Molecule type: mRNA
A:Residues: 1-183 <VAR>
A:Cross-references: EMBL:X14678; NID:g54803; PIDN:CAA32807.1; PID:g54804
C:Genetics:
A:Gene: TIS11
C:Superfamily: Caenorhabditis elegans hypothetical protein Y116A8C.20
C:Keywords: DNA binding; phosphoprotein; zinc finger

Query Match 0.6%; Score 7; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 76;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 ILFKISS 278
|||||||
Db 9 ILFKISS 15

RESULT 44
T15923
hypothetical protein EED8.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15923
R:Chissoe, S.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid EED8.
A:Reference number: Z18428
A:Accession: T15923
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-197 <CHI>
A:Cross-references: EMBL:U23484; NID:g733597; PID:g733609; PIDN:AAC46772.1; CESP:EEDC
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:EED8.12
A:Introns: 103/2

Query Match 0.6%; Score 7; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 TIEVEEE 845
|||||||
Db 73 TIEVEEE 79

RESULT 45
A84865
probable glutathione peroxidase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84865
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84865
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <STO>
A:Cross-references: GB:AE002093; NID:g2289006; PIDN:AAB64335.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g43350
A:Map position: 2
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
|||||||
Db 76 VASKCGL 82

RESULT 46
E75310
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C:Accession: E75310

R.White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1. Science 286, 1571-1577, 1999

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: E75310

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-206 <WHI>

A:Cross-references: GB:AE002048; GB:AE000513; NID:g6459929; PIDN:AAF11690.1; PID:g645993

A:Experimental source: strain R1

C:Genetics:

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DR2136

Query Match 0.6%; Score 7; DB 2; Length 206;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 959 ADMEVDG 965

Db 140 ADMEVDG 146

|||||||

RESULT 47

F83815

hypothetical protein BH1326 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: F83815

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: F83815

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-207 <STO>

A:Cross-references: GB:AF001511; GB:BA000004; NID:g10173727; PIDN:BA05045.1; GSPDB:GNOC

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH1326

C:Superfamily: Bacillus subtilis conserved hypothetical protein yqeJ

Query Match 0.6%; Score 7; DB 2; Length 207;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 GPEVNV 377

Db 152 GPEVNV 158

|||||||

RESULT 48

JC2368

ribosomal protein L13, cytosolic [validated] - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C:Accession: JC2368; PC2232; JC2235; PC2157

R:Olvera, J.; Wool, I.G.

Biochem. Biophys. Res. Commun. 201, 102-107, 1994

A:Title: The primary structure of rat ribosomal protein L13.

A:Reference number: JC2235; MUID:94256964; PMID:8198561

A:Accession: JC2368

A:Molecule type: mRNA

A:Residues: 1-211 <OLV1>

A:Cross-references: EMBL:X78327; NID:g510551; PIDN:CAA55130.1; PID:g510552

A:Accession: PC2232

A:Molecule type: protein

A:Residues: 2-14;26-60;100-180 <OLV2>

A:Experimental source: clone pL13-2,3

A:Note: the protein is designated as ribosomal protein L13

C:Superfamily: rat ribosomal protein L13

C:Keywords: protein biosynthesis; ribosome

F;2-211/Product: ribosomal protein L13 #status predicted <WAT>

Query Match 0.6%; Score 7; DB 1; Length 211;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304

Db 98 VDPRRRN 104

|||||||

RESULT 49

S23753

ribosomal protein L13, cytosolic - human

N:Alternate names: BBL1 protein

C:Species: Homo sapiens (man)

C>Date: 27-May-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999

C:Accession: S23753

R:Adams, S.M.; Helps, N.R.; Sharp, M.G.F.; Brammar, W.J.; Walker, R.A.; Varley, J.M.

Hum. Mol. Genet. 1, 91-96, 1992

A:Title: Isolation and characterization of a novel gene with differential expression

A:Reference number: S23753; MUID:93244791; PMID:1301162

A:Accession: S23753

A:Molecule type: mRNA

A:Residues: 1-211 <ADA>

A:Cross-references: EMBL:X64707; NID:g29382; PIDN:CAA45963.1; PID:g29383

C:Superfamily: rat ribosomal protein L13

C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 0.6%; Score 7; DB 2; Length 211;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304

Db 98 VDPRRRN 104

|||||||

RESULT 50

T22984

hypothetical protein F59B8.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22984

R:Matthews, P.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z19647

A:Accession: T22984

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-213 <WIL>

A:Cross-references: EMBL:Z68343; PIDN:CAA92779.1; GSPDB:GN00022; CESP:F59B8.1

A:Experimental source: clone F59B8

C:Genetics:

A:Gene: CBSP:F59B8.1

A:Map position: 4

A:Introns: 31/2; 65/3; 94/2; 138/1; 170/3

Query Match 0.6%; Score 7; DB 2; Length 213;

Best Local Similarity 100.0%; Pred. No. 87;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 LHDSLKA 641

Db 132 LHDSLKA 138

|||||||

Search completed: November 6, 2002, 03:39:12

Job time : 91 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:26:03 ; Search time 87 Seconds

(without alignments)

531.088 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 1114

Sequence: 1 MGKTIQVGFPPYLLSAEYVK.....RPVNLUSSLRQAQLSHRLVLK 1114

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	%	ID	Description
1	9	0.8	FA5_PIG	Q9g1p1 sus scrofa
2	8	0.7	YDLF_SCHPO	Q10245 schizosacch
3	8	0.7	HIS8_CLOAB	Q97es6 clostridium
4	8	0.7	PBXD_YEAST	P80667 saccharomyc
5	8	0.7	MXK2_YEAST	P32491 saccharomyc
6	8	0.7	DXS_RHOCA	P26242 rhodobacter
7	8	0.7	GREX_BORBU	O31157 borrelia bu
8	8	0.7	PEX1_HUMAN	O43933 homo sapien
9	8	0.7	RRPO_PLAMV	Q07518 plantago as
10	7	0.6	REV_HV128	P05869 human immun
11	7	0.6	RL17_CHLTR	P47760 chlamydia t
12	7	0.6	RL17_CHLMU	Q9pjn5 chlamydia m
13	7	0.6	CYNS_SULTO	Q972w5 sulfolobus
14	7	0.6	GHZ2_HELAN	O33970 helianthus
15	7	0.6	GSX2_ARATH	O04922 arabidopsis
16	7	0.6	BTUE_ECOLI	P06610 escherichia
17	7	0.6	YQOC_CAEEL	Q09301 caenorhabdi
18	7	0.6	NADD_BACHD	Q9kd91 bacillus ha
19	7	0.6	RL13_CHICK	P41125 gallus gall
20	7	0.6	RL13_CRIGR	Q92133 cricetus
21	7	0.6	RL13_HUMAN	P26373 homo sapien
22	7	0.6	RL13_ICTPU	Q90y75 ictalurus p
23	7	0.6	RL13_MOUSE	P47963 mus musculu
24	7	0.6	RL13_RATTU	P41123 rattus norv
25	7	0.6	YDH1_HSVS7	P25049 herpesvirus
26	7	0.6	RL13_SPOPR	Q962u1 spodopora
27	7	0.6	GTC1_RAT	P04904 rattus norv
28	7	0.6	GTC2_RAT	P46418 rattus norv
29	7	0.6	KPYK_THELI	Q56301 thermococcu
30	7	0.6	ETUQ_SALTY	Q9zfv5 salmonele
31	7	0.6	RADC_BACSU	Q02170 bacillus s
32	7	0.6	ENTH1	P42648 entamoeba h
33	7	0.6	CCS_MOUSE	Q9wu84 mus musculu

34	7	0.6	276	1	UPK_BACSU	P94507 bacillus su
35	7	0.6	291	1	Y620_METJA	Q58037 methanococ
36	7	0.6	303	1	UL07_HSVBEB	P28945 equine herp
37	7	0.6	307	1	GLSA_STRCO	P57755 streptomyc
38	7	0.6	311	1	LIP_PSPSP	P26877 pseudomonas
39	7	0.6	314	1	ISPH_BACHD	Q9kd37 bacillus ha
40	7	0.6	320	1	VG2_SPV4	P11334 spiroplasma
41	7	0.6	329	1	Y526_CHLPPN	Q9z826 chlamydia p
42	7	0.6	332	1	PDXA_FUSNN	Q8rgro fusobacteri
43	7	0.6	333	1	RPOA_PSEAE	Q52760 pseudomonas
44	7	0.6	335	1	YO94_CAEEL	P41844 caenorhabdi
45	7	0.6	342	1	RUVE_RICCN	Q92187 rickettsia
46	7	0.6	342	1	RUVE_RICPR	Q9z4e5 rickettsia
47	7	0.6	345	1	NQO8_PARDE	P29920 parascoccus
48	7	0.6	358	1	YJ9L_YEAST	P47175 saccharomyc
49	7	0.6	362	1	BIOB_SYNY3	P73538 synchocyst
50	7	0.6	371	1	YPC4_CAEEL	Q11181 caenorhabdi

ALIGNMENTS

RESULT 1
FA5_PIG ID FA5_PIG STANDARD; PRT; 2358 AA.
AC Q9GLP1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID:9623;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C DOMAINS.
RC TISSUE=Liver;
RX MEDLINE=21121490; PubMed=11229814;
RA Grimm H.K.W.; Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional protein modeling of membrane binding sites and comparative anatomy of domains.";
RL Cell. Mol. Life Sci. 58:148-159(2001).
CC -!- FUNCTION: Coagulation factor V is a cofactor that participates with factor Xa to activate prothrombin to thrombin.
CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light chain, noncovalently bound. The interaction between the two chains is calcium-dependent.
CC -!- DOMAIN: Domain B contains 41 x 9 AA tandem repeats. Domains C1 and C2 may be involved in membrane binding.
CC -!- PM: Thrombin activates factor V proteolytically to the active cofactor, factor Va (formation of a heavy chain at the N-terminus and a light chain at the C-terminus).
CC -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF 2 PLASTOCYANIN-LIKE REPEATS.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; AF191308; AAC28381.1; -
HSSP; P12259; ICZT.
InterPro; IPR001117; Cu-oxidase.
InterPro; IPR000421; FA58_C.

RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgeson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovsky G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of *Schizosaccharomyces pombe*.";
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. STRONG, TO YEAST YBR159W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z69727; CAA93565.1; .
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00166; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
FT ACT_SITE 216 216 BY SIMILARITY.
SQ SEQUENCE 341 AA; 37307 MW; F49CAAB79486F71D CRC64;

Query Match 0.7%; Score 8; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1025 RALRKEAR 1032
DB 327 RALRKEAR 334
|||||||

RESULT 3
HIS8_CLOB STANDARD; PRT; 352 AA.
AC Q97BS6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol-
DE phosphate transaminase).
GN HIS8 OR CAC3031.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium *Clostridium acetobutylicum*.";
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate = 3-
CC (imidasol-4-yl)-2-oxopropyl phosphate + L-glutamate.
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- PATHWAY: Histidine biosynthesis; seventh step.
CC -!- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES. HISTIDINOL-PHOSPHATE AMINOTRANSFERASES
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB007800; AAK80971.1; .
DR InterPro: IPR004839; Aminotransf1/2.
DR InterPro: IPR001917; NHtransf_2.
DR Pfam: PF00155; aminotran_1.2; 1.
DR TIGRFAMS: TIGR01141; hisc; 1.
DR PROSITE: PS00599; AA_TRANSFER_CLASS_2; 1.
KW Histidine biosynthesis; Transferase; Aminotransferase;
KW Pyridoxal phosphate; Complete proteome.
FT BINDING 210 210 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 352 AA; 40220 MW; 3375021EBEC6BC2E CRC64;

Query Match 0.7%; Score 8; DB 1; Length 352;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 IDNFLRVS 395
DB 326 IDNFLRVS 333
|||||||

RESULT 4
PEXD_YEAST STANDARD; PRT; 386 AA.
AC P80667;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Peroxisomal membrane protein PAS20 (Peroxin-13).
GN PEX13 OR PAS20 OR YLR191W OR L9470.1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97011156; PubMed=8858166;
RA Elgersma Y., Kwast L., Klein A., Voorn-Brouwer T., van den Berg M.,
RA Tabak H.F., Distel B.;
RT "The SH3 domain of the Saccharomyces cerevisiae peroxisomal membrane
RT protein Pex13p functions as a docking site for Pex5p, a mobile
RT receptor for the import PTS1-containing proteins.";
RL J. Cell Biol. 135:97-109(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,

Query Match 0.7%; Score 8; DB 1; Length 506;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 923 ISERVIGK 930
| | | | |
DB 314 ISERVIGK 321

RESULT 6

DXS_RHOCA STANDARD; PRT; 641 AA.
AC P26242;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1-deoxy-D-xylulose 5-phosphate synthase (EC 4.1.3.37) (1-
DE deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).
GN DXS.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84259352; PubMed=6744416;
RA Youvan D.C., Bylina E.J., Alberti M., Begusch H., Hearst J.E.;
RT "Nucleotide and deduced polypeptide sequences of the photosynthetic
RT reaction-center, B870 antenna, and flanking polypeptides from R.
RT capsulata".
RL Cell 37:949-957(1984).
CC -1- FUNCTION: Catalyzes the acyloln condensation reaction between C
CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
CC 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).
CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
CC deoxy-D-xylulose 5-phosphate + CO(2).
CC -1- COFACTOR: Thiamine pyrophosphate (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.
CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
CC step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY. DXS SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K01183; -; NOT ANNOTATED_CDS.
CC EMBL; X11165; CAA77557.1; -.
CC PIR; G28771; G28771.
CC InterPro; IPR000360; Transketolase.
CC Pfam; PF02779; transket_pyr; 1.
CC Pfam; PF02780; transketolase_C; 1.
CC TIGRfam; TIGR00204; dxs; 1.
CC PROSITE; PS00801; TRANSKETOLASE_1; 1.
CC PROSITE; PS00802; TRANSKETOLASE_2; 1.
CC Lysase; Flavo-protein; Thiamine pyrophosphate;
KW Isoprene biosynthesis; Thiamine biosynthesis.
SQ SEQUENCE 641 AA; 67943 MW; CBCC0BF23C012201 CRC64;

Query Match 0.7%; Score 8; DB 1; Length 641;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 GLTANDIR 472
| | | | |
DB 615 GLTANDIR 622

RESULT 7

GREB_BORBU STANDARD; PRT; 901 AA.
AC O51157;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription elongation factor greA (Transcript cleavage factor
DE greA).
GN GREB OR BB0132.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kervlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi".
RL Nature 390:580-586(1997).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION
CC ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING
CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION
CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN
CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY
CC CLEAVAGE FACTORS SUCH AS GREB OR GREB ALLOWS THE RESUMPTION OF
CC ELONGATION FROM THE NEW 3' TERMINUS. GREB RELEASES SEQUENCES OF
CC 2 TO 3 NUCLEOTIDES (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 GREB ASSOCIATED DOMAIN 1 (GRAD1).
CC -1- SIMILARITY: CONTAINS 1 GREB ASSOCIATED DOMAIN 2 (GRAD2).
CC -1- SIMILARITY: BELONGS TO THE GREB/GREB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001124; AAC66506.1; -.
CC HSSP; P21346; IGRJ.
CC TIGR; BB0132; -.
CC InterPro; IPR001437; GreA_Greb.
CC Pfam; PF01272; GreA_Greb; 1.
CC Pfam; PF03449; GreA_Greb_N; 1.
CC ProDom; PD004918; GreA_Greb; 1.
CC PROSITE; PS00829; GREAB_1; 1.
CC PROSITE; PS00830; GREAB_2; 1.
CC Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 20 170 GRAD1.
FT DOMAIN 171 681
FT DOMAIN 741 901 GREB.
FT DOMAIN 784 815 COILED COIL (POTENTIAL).
SQ SEQUENCE 901 AA; 105314 MW; FBB81088D735A5F9 CRC64;

Query Match 0.7%; Score 8; DB 1; Length 901;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 764 DIRVLKAV 771
| | | | |
DB 359 DIRVLKAV 366

```
RESULT 8
PEX1_HUMAN
ID PEX1_HUMAN STANDARD; PRT; 1283 AA.
AC O43933; Q99994;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxisome biogenesis factor 1 (peroxin-1) (Peroxisome biogenesis
DE disorder protein 1).
GN PEX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT NALD ASP-843.
RX MEDLINE=98061095; PubMed=9398848;
RA Portsteffen H., Beyer A., Becker E., Epplen C., Pawlak A.,
RA Kunau W.-H., Dödt G.;
RT "Human PEX1 is mutated in complementation group 1 of the peroxisome
RT biogenesis disorders";
RL Nat. Genet. 17:449-452(1997).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT NALD ASP-843.
RX MEDLINE=98061094; PubMed=9398847;
RA Reuber B.E., Germain-Lee E., Collins C.S., Morrell J.C.,
RA Ameritunga R., Moser H.W., Valle D., Gould S.J.;
RT "Mutations in PEX1 are the most common cause of peroxisome biogenesis
RT disorders";
RL Nat. Genet. 17:445-448(1997).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS NALD PRO-664 AND ASP-843.
RX MEDLINE=98208543; PubMed=9539740;
RA Tamura S., Okumoto K., Toyama R., Shimozawa N., Tsukamoto T.,
RA Suzuki Y., Osumi T., Kondo N., Fujiki Y.;
RT "Human PEX1 cloned by functional complementation on a CHO cell mutant
RT is responsible for peroxisome-deficient Zellweger syndrome of
RT complementation group I.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:4350-4355(1998).
RN [4]
RP SEQUENCE OF 635-1283 FROM N.A.
RA Pauley A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR STABILITY OF PEX5 AND PROTEIN IMPORT INTO
CC THE PEROXISOME MATRIX.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- DISEASE: DEFECTS IN PEX1 ARE THE CAUSE OF PEROXISOME BIOGENESIS
CC DISORDERS (PBD) WHICH ARE LETHAL RECESSIVE DISEASES CAUSED BY
CC DEFECTS IN PEROXISOME ASSEMBLY. NEONATAL ADRENOLKODYSTROPHY
CC (NALD) IS ONE OF THE PBD LINKED TO DEFECTS IN PEX1
CC -!- DISEASE: DEFECTS IN PEX1 ARE THE CAUSE OF INFANTILE REFSUM DISEASE
CC (IRD). IRD IS A FORM OF PHYTANIC ACID STORAGE DISEASE.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF030356; AAB99758.1; -
CC DR EMBL; AF026086; AAB87880.1; -
CC DR EMBL; AB008112; BA885162.1; -
CC DR EMBL; AC000084; ABA846346.1; ALT_SEQ.
CC DR Genew; HGNC:8850; PEX1.
CC MIM; 602136; -
CC DR MIM; 202370; -
CC DR MIM; 266510; -
CC DR InterPro; IPR003593; AAA_ATPase.
CC DR InterPro; IPR003959; AAA_ATPase_cent.
CC DR InterPro; IPR003960; AAA_sub.

RESULT 9
RRPO_PIAMV
ID RRPO_PIAMV STANDARD; PRT; 1385 AA.
AC Q07518;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA replication protein (156 kDa protein) (ORF 1) [Contains: RNA-
DE directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Plantago asiatica mosaic potexvirus (PIAMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
OX NCBI_TaxID=45416;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94157481; PubMed=8113746;
RA Solov'yev A.G., Novikov V.K., Merits A., Savenkov E.I.,
RA Zelenina D.A., Tyukina L.G., Morozov S.Y.;
RT "Genome characterization and taxonomy of Plantago asiatica mosaic
RT potexvirus.";
RL J. Gen. Virol. 75:259-267(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93237851; PubMed=8477237;
RA Solov'yev A.G., Novikov V.K., Morozov S.I., Kagramanov V.N.,
RA Atabekov I.G.;
RT "Primary structure of the triple block RNA genes of the Plantago
RT asiatica mosaic virus.";
RL Dokl. Akad. Nauk SSSR 328:625-628(1993).
CC -!- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z21647; CAA79761.1; -
CC DR EMBL; IPR001788; RNA_dep_RNAPol2.
CC DR InterPro; IPR000606; Viral_helicase1.
CC DR Fram; PF00978; RNA_dep_RNAPol2; 1.
CC DR Fram; PF01443; Viral_helicase1; 1.
CC DR ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
CC Transferase.
CC NP_BIND 665 672 ATP (POTENTIAL).
CC SEQUENCE 1385 AA; 155616 MW; 56C177E44678A94F CRC64;

Query Match 0.7%; Score 8; DB 1; Length 1385;
Best Local Similarity 100.0%; Pred. No. 13; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 68 RLYFGSSY 75
Db 1127 RLYFGSSY 1134
|||||||
Db 1127 RLYFGSSY 1134

RESULT 9
RRPO_PIAMV
ID RRPO_PIAMV STANDARD; PRT; 1385 AA.
AC Q07518;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA replication protein (156 kDa protein) (ORF 1) [Contains: RNA-
DE directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Plantago asiatica mosaic potexvirus (PIAMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
OX NCBI_TaxID=45416;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94157481; PubMed=8113746;
RA Solov'yev A.G., Novikov V.K., Merits A., Savenkov E.I.,
RA Zelenina D.A., Tyukina L.G., Morozov S.Y.;
RT "Genome characterization and taxonomy of Plantago asiatica mosaic
RT potexvirus.";
RL J. Gen. Virol. 75:259-267(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93237851; PubMed=8477237;
RA Solov'yev A.G., Novikov V.K., Morozov S.I., Kagramanov V.N.,
RA Atabekov I.G.;
RT "Primary structure of the triple block RNA genes of the Plantago
RT asiatica mosaic virus.";
RL Dokl. Akad. Nauk SSSR 328:625-628(1993).
CC -!- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z21647; CAA79761.1; -
CC DR EMBL; IPR001788; RNA_dep_RNAPol2.
CC DR InterPro; IPR000606; Viral_helicase1.
CC DR Fram; PF00978; RNA_dep_RNAPol2; 1.
CC DR Fram; PF01443; Viral_helicase1; 1.
CC DR ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
CC Transferase.
CC NP_BIND 665 672 ATP (POTENTIAL).
CC SEQUENCE 1385 AA; 155616 MW; 56C177E44678A94F CRC64;

Query Match 0.7%; Score 8; DB 1; Length 1385;
Best Local Similarity 100.0%; Pred. No. 13; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 68 RLYFGSSY 75
Db 1127 RLYFGSSY 1134
|||||||
Db 1127 RLYFGSSY 1134
```

```
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 ERILSTLR 432
Db 597 ERILSTLR 604

RESULT 10
REV_HV128 STANDARD; PRT; 90 AA.
ID REV_HV128
AC P05869;
DT 01-NOV-1988 (Rel. 09, Created)
DI 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE REV protein (Anti-repression transactivator protein) (ART/TRS)
DE (Fragment).
GN REV.
OS Human immunodeficiency virus type 1 (Z-84 isolate) (HIV-1).
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11681;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88281278; PubMed=3395517;
RA Yourho J., Josephs S.F., Reitz M.S. Jr., Zagury D., Wong-Staal F.,
RA Gallo R.C.;
RT "Nucleotide sequence analysis of the env gene of a new Zairian
RT isolate of HIV-1.";
RL AIDS Res. Hum. Retroviruses 4:165-173(1988).
CC -!- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
CC -!- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
CC -!- PTM: PHOSPHOPROTEIN WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
CC BY A SPECIFIC SERINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
CC -!- MISCELLANEOUS: THE Z-84 ISOLATE WAS TAKEN FROM A 54 YEAR-OLD
CC ZAIREAN MALE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03653; AAA44686.1; -
DR HIV; J03653; REV5JY1.
DR InterPro; IPR000625; REV_protein.
DR Pfam; PF00424; REV; 1
KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
FT NON_TER 1
SQ SEQUENCE 90 AA; 10031 MW; B1940D344B0580B7 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 90;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 ERILSTL 431
Db 31 ERILSTL 37

RESULT 11
RL17_CHLTR STANDARD; PRT; 141 AA.
ID RL17_CHLTR
AC P47760; O84514;
DT 01-FEB-1996 (Rel. 33, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L17.
GN RPLQ OR RL17 OR CT506.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
```

```
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=L2/434/Bu; PubMed=7730299;
RX MEDLINE=95247702;
RA Gu L.J., Wenman W.M., Remacha M., Meuser R.U., Coffin J.M., Kaul R.;
RT "Chlamydia trachomatis RNA polymerase alpha subunit: sequence and
RT structural analysis.";
RL J. Bacteriol. 177:2594-2601(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=D/UW-3/Cx; PubMed=97841136;
RX MEDLINE=99000809; PubMed=97841136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT WAS CORRECTED IN POSITION 14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L39334; AAA74990.1; ALT_INIT.
DR EMBL; AE01323; AAC68107.1; -
DR InterPro; IPR000456; Ribosomal_L17.
DR Pfam; PF01196; Ribosomal_L17; 1
DR ProDom; PD004277; Ribosomal_L17; 1
DR TIGRfams; TIGR00059; L17; 1.
DR PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW Ribosomal protein; Complete proteome.
FT VARIANT 116 116 G->R (IN SEROVAR L2).
SQ SEQUENCE 141 AA; 16152 MW; DAD293A8FBE8D79 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 RVIGKLF 932
Db 98 RVIGKLF 104

RESULT 12
RL17_CHLMU STANDARD; PRT; 142 AA.
ID RL17_CHLMU
AC Q9PJN5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L17.
GN RPLQ OR TC0793.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MoPn / Nigg; PubMed=10684935;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
```


RT pneumoniac AR39"; 28:1397-1406(2000).
RL Nucleic Acids Res.
CC -!- SIMILARITY: BELONGS TO THE LI7P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002347; AAF39596.1; .
DR TIGR; TC0793; .
DR InterPro; IPR000456; Ribosomal_L17.
DR Pfam; PF01196; Ribosomal_L17; 1.
DR ProDom; PD004277; Ribosomal_L17; 1.
DR TIGRFAMs; TIGR00059; L17; 1.
DR PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 142 AA; 16223 MW; B17AD1A1760A0767 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 142;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 RVIGKLF 932
|||||||
DB 98 RVIGKLF 104

RESULT 13
CYNS_SULTO
ID CYNS_SULTO STANDARD; PRT; 148 AA.
AC Q972W5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cyanate hydratase (EC 4.2.1.104) (Cyanase) (Cyanate lyase) (Cyanate
DE hydratase).
GN CYNS OR STI023.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S., Y.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermocacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
CC -!- FUNCTION: Catalyzes the reaction of cyanate with bicarbonate to
CC produce ammonia and carbon dioxide (by similarity).
CC -!- CATALYTIC ACTIVITY: Cyanate (NCO(-)) + H(2)O = carbamate (H(2)N-
CC CO-O(-)).
CC
CC -!- SIMILARITY: BELONGS TO THE CYANASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000984; BAB66048.1; .

DR InterPro; IPR003712; Cyanate_lyase.
DR Pfam; PF02560; Cyanate_lyase; 1.
DR TIGRFAMs; TIGR00673; cynS; 1.
KW Lyase; Complete proteome.
FT ACT_SITE 89 89 BY SIMILARITY.
FT ACT_SITE 92 92 BY SIMILARITY.
FT ACT_SITE 115 115 BY SIMILARITY.
SQ SEQUENCE 148 AA; 17410 MW; 33CB637A45B3BB85 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ADKIITL 64
|||||||
DB 51 ADKIITL 57

RESULT 14
GSHZ_HELAN
ID GSHZ_HELAN STANDARD; PRT; 167 AA.
AC Q23970;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glutathione peroxidase 1 (EC 1.11.1.9).
GN GPXHA-1.
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
RN [1]
RP SEQUENCE FROM N.A.
RA Drevet J.R., Gagne G., Tourvieille de Labrouhe D., Nicolas P.,
RA Dufau J.P., Ledoigt G., Roedel-Drevet P.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY CONSTITUTE A GLUTATHIONE PEROXIDASE-LIKE
CC PROTECTIVE SYSTEM AGAINST OXIDATIVE STRESSES.
CC -!- CATALYTIC ACTIVITY: 2 glutathione + H(2)O(2) = oxidized
CC glutathione + 2 H(2)O.
CC
CC -!- SIMILARITY: BELONGS TO THE GLUTATHIONE PEROXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y14429; CAA74775.1; .
DR HSSP; P00435; LGPI.
DR InterPro; IPR000889; Glut_peroxidase.
DR Pfam; PF00255; GSHPx; 1.
DR PRINTS; PR01011; GLUTPROXDASE.
DR PROSITE; PS00460; GLUTATHIONE_PEROXID_1; 1.
DR PROSITE; PS00763; GLUTATHIONE_PEROXID_2; 1.
KW Oxidoreductase; Peroxidase.
FT ACT_SITE 41 41
SQ SEQUENCE 167 AA; 18842 MW; CFSAC/ACC76558C9 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 167;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545
|||||||
DB 37 VASKCGL 43

RESULT 15

GSX ARATH
ID GSHX_ARATH STANDARD: PRT: 169 AA.
AC 004922;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable glutathione peroxidase At2g51570 (EC 1.11.1.9).
GN AT2G31570 OR T9H9.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eudicots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Turano F.J., Caldwell C.R., McMahon M.;
RT "Glutathione peroxidase cDNA from Arabidopsis";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowmar C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H.B.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Coppenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana";
RL Nature 402:761-768(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY CONSTITUTE A GLUTATHIONE PEROXIDASE-LIKE
PROTECTIVE SYSTEM AGAINST OXIDATIVE STRESSES.
CC -!- CATALYTIC ACTIVITY: 2 glutathione + H(2)O(2) = oxidized
glutathione + 2 H(2)O.
CC -!- SIMILARITY: BELONGS TO THE GLUTATHIONE PEROXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U94495; AAB52725.1; -
DR EMBL; AC007071; AAD24836.1; -
DR EMBL; AY058187; AAL25600.1; -
DR EMBL; AY044330; AAK73271.1; -
DR HSP; P00435; 1GPI.
DR InterPro; IPR000889; Glut_peroxidase.
DR Pfam; PF00255; GSHPX; 1.
DR PRINTS; PR01011; GLUTPROXASE.
DR PROSITE; PS00460; GLUTATHIONE_PEROXID_1; 1.
DR PROSITE; PS00763; GLUTATHIONE_PEROXID_2; 1.
KW Oxidoreductase; Peroxidase.
FT ACT_SITE 41 BY SIMILARITY.

SQ SEQUENCE 169 AA; 18944 MW; 23930B83A0AE3251 CRC64;
Query Match 0.6%; Score 7; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 539 VASKCGL 545
DB 37 VASKCGL 43
|||||
|||
RESULT 16
BTUE_ECOLI
ID BTUE_ECOLI STANDARD: PRT: 183 AA.
AC P06610;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vitamin B12 transport periplasmic protein btue.
GN BTUE OR B1710.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304184; PubMed=3528129;
RA Friedrich M.J., Deveau L.C., Kadner R.J.;
RT "Nucleotide sequence of the btueD genes involved in vitamin B12
transport in Escherichia coli and homology with components of
periplasmic-binding-protein-dependent transport systems";
RL J. Bacteriol. 167:928-934(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1232-1244(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kimura K., Kimura S., Kitakawa M., Mori T.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori Y.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubramanian S.,
RA Tagami H., Takada J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map";
RL DNA Res. 3:363-377(1996).
CC -!- FUNCTION: NOT ESSENTIAL FOR B12 TRANSPORT; HOWEVER, IT IS AN
AUXILIARY COMPONENT OF THE TRANSPORT SYSTEM.
CC -!- SUBCELLULAR LOCATION: Periplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE GLUTATHIONE PEROXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M14031; AAA33527.1; -
DR EMBL; AE000266; AAC74780.1; -
DR EMBL; D90813; BAA15478.1; -
DR EMBL; D90814; BAA15490.1; -

```

DR PIR; E24496; QRECEB.
DR HSSP; P00435; 1GP1.
DR EcoGene; EG10129; btuE.
DR InterPro; IPR000889; Glut_peroxidase.
DR Pfam; PF00255; GSHPx; 1.
DR PRINTS; PR01011; GLUTPOXDASE.
DR PROSITE; PS00460; GLUTATHIONE_PEROXID_1; 1.
DR PROSITE; PS00763; GLUTATHIONE_PEROXID_2; 1.
DR Petiplasmic; Cobalt transport; Complete proteome.
KW PERiplasmic; Cobalt transport; Complete proteome.
SQ SEQUENCE 183 AA; 20469 MW; C8DB671963A7F235 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 539 VASKGGL 545
Db 33 VASKGGL 39

RESULT 17
YQOC_CAEEL STANDARD; PRT; 137 AA.
AC Q09301;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical RNA-binding protein EED8.12 in chromosome II.
GN EED8.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Chisoe S.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -!- SIMILARITY: STRONG, TO C.ELEGANS EED8.4.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U23484; AAC46772.1; -.
DR WormPep; EED8.12; CE01885.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW Hypothetical protein; RNA-binding.
FT DOMAIN 51 138 RNA-BINDING (RRM).
SQ SEQUENCE 137 AA; 21649 MW; AB5C4E7247DD9729 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 839 TIEEVEE 845
Db 73 TIEEVEE 79

RESULT 18
NADD_BACHD STANDARD; PRT; 207 AA.
ID NADD_BACHD
AC Q9K9D1;

```

```

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable nicotinate-nucleotide adenyltransferase (EC 2.7.7.18)
DE (deamido-NAD(+)-pyrophosphorylase) (deamido-NAD(+) diphosphorylase)
DE (Nicotinate mononucleotide adenyltransferase) (NAMN
DE adenyltransferase).
DE NADD OR BH1326.
GN Bacillus halodurans.
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: CATALYZES THE REVERSIBLE ADENYLATION OF NICOTINATE
CC MONONUCLEOTIDE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + nicotinate ribonucleotide - diphosphate
CC + deamido-NAD(+).
CC -!- PATHWAY: NAD biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE NADD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001511; BAB05045.1; -.
DR InterPro; IPR004821; Cyt_tran_rel.
DR InterPro; IPR004820; Cytidylyltransf.
DR InterPro; IPR005248; NAMN_adntrnsfrase.
DR Pfam; PF01467; Cytidylyltransf; 1.
DR TIGRFAMs; TIGR00125; cyt_tran_rel; 1.
DR TIGRFAMs; TIGR00482; TIGR00482; 1.
KW Transferase; Nucleotidytransferase; NAD; Complete proteome.
SQ SEQUENCE 207 AA; 23434 MW; C2F56B3F85109632 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 371 GPEVNVNS 377
Db 152 GPEVNVNS 158

RESULT 19
RLI3_CHICK STANDARD; PRT; 210 AA.
ID RLI3_CHICK
AC P41125;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L13 (Breast basic conserved protein 1).
GN RPL13 OR BBL1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens fibers;
RX MEDLINE=96437509; PubMed=8840185;

```

```
RA Savada K., Agata K., Eguchi G.;
RT "Characterization of terminally differentiated cell state by
RT categorizing cDNA clones derived from chicken lens fibers.";
RL Int. J. Dev. Biol. 40:531-535(1996).
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D26318; BAA05377.1; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT_MET 0
SQ SEQUENCE 210 AA; 24264 MW; D8D005B96A8BA9B4 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304
DB 97 VDPRRRN 103
|||||

RESULT 20
RL13_CRIGR
ID RL13_CRIGR STANDARD; PRT; 210 AA.
AC Q92313;
AT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 60S ribosomal protein L13.
GN RPL13.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki Y., Itoh F., Suzuki H., Hinoda Y., Imai K.;
RT "Identification of genes highly expressed in association with G2
RT arrest induced by DNA damage of Chinese hamster ovary cells by
RT differential display.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014876; BAA34291.1; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT_MET 0
SQ SEQUENCE 210 AA; 24261 MW; 484429F34C565720 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 210;
```

```
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304
DB 97 VDPRRRN 103
|||||

RESULT 21
RL13_HUMAN
ID RL13_HUMAN STANDARD; PRT; 210 AA.
AC P26373; Q9BPX0;
AT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 60S ribosomal protein L13 (Breast basic conserved protein 1).
GN RPL13 OR BEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Adams S.M., Helps N.R., Sharp M.G.F., Brammar W.J., Walker R.A.,
RA Varley J.M.;
RT "Isolation and characterization of a novel gene with differential
RT expression in benign and malignant human breast tumours.";
RL Hum. Mol. Genet. 1:91-96(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cervix, Lung, Lymph, Pancreas, Placenta, and Uterus;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- TISSUE SPECIFICITY: HIGHER LEVELS OF EXPRESSION IN BENIGN BREAST
CC LESIONS THAN IN CARCINOMAS.
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64707; CAA45963.1; -
DR EMBL; BC004954; RAH04954.1; -
DR EMBL; BC007345; RAH07345.1; -
DR EMBL; BC007563; RAH07563.1; -
DR EMBL; BC007805; RAH07805.1; -
DR EMBL; BC010994; RAH10994.1; -
DR EMBL; BC013078; RAH13078.1; -
DR EMBL; BC014167; RAH14167.1; -
DR EMBL; BC020804; RAH20804.1; -
DR EMBL; BC027463; RAH27463.1; -
DR PIR; S23753; S23753
DR Genew; HGNC:10303; RPL13.
DR MIM; 113703; -
DR InterPro; IPR001380; Ribosomal_L13e.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT_MET 0
SQ SEQUENCE 210 AA; 24130 MW; CBBB26F143FAF058 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304
```

```

Db 97 VDPERRN 103
|||||
RESULT 22
RL13_ICTPU STANDARD; PRT; 210 AA.
AC Q90YV5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L13.
GN RPL13
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ictalurus.
OX NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RA Patterson A.P., Karsi A., Liu Z.J.;
RT "Translational machinery of channel catfish: II. Complementary DNA and
RT expression of the complete set of 47 60S ribosomal proteins."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF401567; AAK95139.1;
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT MET 0 BY SIMILARITY.
FT CONFLICT 39 39 Q -> L (IN REF. 1).
FT CONFLICT 189 195 RLFGR -> PTLWQSEQ (IN REF. 1).
FT CONFLICT 202 210 AEQDKKK -> SEQRCWKKRN (IN REF. 1).
SQ SEQUENCE 210 AA; 24254 MW; 5D97B80DD4E64172 CRC64;
Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 298 VDPERRN 304
Db 97 VDPERRN 103
|||||
RESULT 23
RL13_MOUSE STANDARD; PRT; 210 AA.
AC P47963; Q9DCH1; Q9CRZ9;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L13 (A52).
GN RPL13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Palacios R., Xie X.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic liver, and kidney;

```

```

RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang X.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U28917; AAA69923.1;
DR EMBL; AK002787; BAB23358.1;
DR EMBL; AK010989; BAB27309.1;
DR MGD; MGI:105922; Rpl13.
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT MET 0 BY SIMILARITY.
FT CONFLICT 39 39 Q -> L (IN REF. 1).
FT CONFLICT 189 195 RLFGR -> PTLWQSEQ (IN REF. 1).
FT CONFLICT 202 210 AEQDKKK -> SEQRCWKKRN (IN REF. 1).
SQ SEQUENCE 210 AA; 24174 MW; 3795373F518655F2 CRC64;
Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 298 VDPERRN 304
Db 97 VDPERRN 103
|||||
RESULT 24
RL13_RAT STANDARD; PRT; 210 AA.
AC P41123;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 60S ribosomal protein L13.
GN RPL13.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=94256964; PubMed=8198561;
RA Olivera J., Wool I.G.;

```

RT The primary structure of rat ribosomal protein L13.";
RL Biochem. Biophys. Res. Commun. 201:102-107(1994).
CC -!- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X78327; CAA55130.1; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR ProSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT_MET 0
FT CONFLICT 55 55 R -> S (IN CDNA).
SQ SEQUENCE 210 AA; 24178 MW; 379AG737F30C6E9 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304
DB 97 VDPRRRN 103
|||||

RESULT 25
YDHL_HSV57
ID YDHL_HSV57 STANDARD; PRT; 214 AA.
AC P25049.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 24.1 kDa protein in DHFR 3'region (ORF2).
OS Herpesvirus saimiri (strain 484-77).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90279084; PubMed=2161952;
RA Geck P., Whitaker S.A., Medveczky M.M., Medveczky P.G.;
RT "Expression of collagenlike sequences by a tumor virus, herpesvirus
RL J. Virol. 64:3509-3515(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31964; AAA46153.1; -
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 214 AA; 24142 MW; 9869BDBDA89FA42A CRC64;

Query Match 0.6%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 KSFLEYX 26
DB 107 KSFLEYX 113
|||||

RESULT 26

RL13_SPOFR
ID RL13_SPOFR STANDARD; PRT; 219 AA.
AC Q962U1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L13.
GN RPL13.
OS Spodoptera frugiperda (Fall armyworm).
CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
CC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
CC Ditrysia; Noctuoidea; Noctuidae; Amphipyridae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RA Landais I., Ogilastro M., Mita K., Nohata J., Lopez-Ferber M.,
RA Donor-Cerutti M., Fournier P., Devauchelle G.;
RT "Full-length ribosomal protein sequence from an EST library of
RT Spodoptera frugiperda cells (Sf9).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF400183; AAK92155.1; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR ProSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
SQ SEQUENCE 219 AA; 24963 MW; 9FF918EB07CA3449 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304
DB 97 VDPRRRN 103
|||||

RESULT 27
GTCL_RAT
ID GTCL_RAT STANDARD; PRT; 220 AA.
AC P04904;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutathione S-transferase Yc-1 (EC 2.5.1.18) (Chain 2) (GST Yc1)
DE (GST class-alpha).
GN GSTY1.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer 344; TISSUE=Liver;
RX MEDLINE=94327654; PubMed=8051171;
RA Hayes J.D., Nguyen T., Judah D.J., Petersson D.G., Neal G.E.;
RT "Cloning of cDNAs from fetal rat liver encoding glutathione S-
RT transferase Yc polypeptides. The Yc2 subunit is expressed in adult
RT rat liver resistant to the hepatocarcinogen aflatoxin B1.";
RL J. Biol. Chem. 269:20707-20717(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85182743; PubMed=2985614;

RA Telakowski-Hopkins C.A., Rodkey K.A., Bennett C.D., Lu A.Y.H.,
 RA Pickett C.B.;
 RT "Rat liver glutathione S-transferases. Construction of a cDNA clone
 RT complementary to a Yc mRNA and prediction of the complete amino acid
 RT sequence of a Yc subunit.";
 RL J. Biol. Chem. 260:5820-5825(1985).
 FN [3]
 FN SEQUENCE OF 74-220 FROM N.A.
 RP MEDLINE=84264589; PubMed=6204982;
 RX Tu C.-P.D., Lai H.-C.J., Li N.-O., Weiss M.J., Reddy C.C.;
 RA "The Yc and Ya subunits of rat liver glutathione S-transferases are
 RT the products of separate genes.";
 RL J. Biol. Chem. 259:9434-9439(1984).
 FN [4]
 FN PARTIAL SEQUENCE.
 RP STRAIN=Fischer 344; TISSUE=Liver;
 RX MEDLINE=92061940; PubMed=1953636;
 RA Hayes J.D., Judah D.J., McLellan L.I., Kerr L.A., Peacock S.D.,
 RA Neal G.E.;
 RT "Ethoxyquin-induced resistance to aflatoxin B1 in the rat is
 RT associated with the expression of a novel alpha-class glutathione
 RT S-transferase subunit, Yc2, which possesses high catalytic activity
 RT for aflatoxin B1-8,9-epoxide.";
 RL Biochem. J. 279:385-398(1991).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES. HAS
 CC SUBSTANTIAL ACTIVITY TOWARD AFLATOXIN B1-8,9-EPOXIDE.
 CC -!- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
 CC -!- SUBUNIT: HETERODIMER OF YC1 AND YC2.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- DEVELOPMENTAL STAGE: LIVER FROM ADULT FEMALE RATS CONTAINS ABOUT
 CC 2-FOLD GREATER LEVELS OF YC1 THAN IS FOUND IN LIVER FROM ADULT
 CC MALE RATS.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X78848; CAA55405.1; -;
 DR EMBL; K01932; AAA41294.1; -;
 DR EMBL; S72505; -; NOT_ANNOTATED_CDS.
 DR PIR; A26753; A26753.
 DR PIR; A26754; A26754.
 DR HSSP; P08263; IGSD.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR004045; GST_Nterm.
 DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF02798; GST_N; 1.
 KW Transferase; Multigene family.
 FT INIT_MET 0
 FT CONFLICT 101 101 L -> I (IN REF. 2).
 FT CONFLICT 183 183 L -> K (IN REF. 2).
 SQ SEQUENCE 220 AA; 25188 MW; F40F5FC0757EE93F CRC64;
 Query Match 0.6%; Score 7; DB 1; Length 220;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1089 KDKARNR 1095
 Db 124 KDKARNR 130
 RESULT 28
 GTC2_RAT
 ID GTC2_RAT STANDARD; PRT; 220 AA.
 AC P46418;
 DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glutathione S-transferase Yc-2 (EC 2.5.1.18) (Chain 2) (GST Yc2)
 DE (GST class-alpha).
 GN (GST Yc2 OR GSTA5).
 DE
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Fischer 344; TISSUE=Liver;
 RX MEDLINE=94327654; PubMed=8051171;
 RA Hayes J.D., Nguyen T., Judah D.J., Petersson D.G., Neal G.E.;
 RT "Cloning of cDNAs from fetal rat liver encoding glutathione S-
 RT transferase Yc polypeptides. The Yc2 subunit is expressed in adult
 RT rat liver resistant to the hepatocarcinogen aflatoxin B1.";
 RL J. Biol. Chem. 269:20707-20717(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=96358489; PubMed=8761455;
 RX Pulford D.J., Hayes J.D.;
 RA "Characterization of the rat glutathione S-transferase Yc2 subunit
 RT gene, GSTA5: identification of a putative antioxidant-responsive
 RT element in the 5'-flanking region of rat GSTA5 that may mediate
 RT chemoprotection against aflatoxin B1.";
 RL Biochem. J. 318:75-84(1996).
 RN [3]
 RN PARTIAL SEQUENCE.
 RP STRAIN=Fischer 344; TISSUE=Liver;
 RX MEDLINE=92061940; PubMed=1953636;
 RA Hayes J.D., Judah D.J., McLellan L.I., Kerr L.A., Peacock S.D.,
 RA Neal G.E.;
 RT "Ethoxyquin-induced resistance to aflatoxin B1 in the rat is
 RT associated with the expression of a novel alpha-class glutathione
 RT S-transferase subunit, Yc2, which possesses high catalytic activity
 RT for aflatoxin B1-8,9-epoxide.";
 RL Biochem. J. 279:385-398(1991).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES. HAS
 CC SUBSTANTIAL ACTIVITY TOWARD AFLATOXIN B1-8,9-EPOXIDE.
 CC -!- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
 CC -!- SUBUNIT: HETERODIMER OF YC1 AND YC2.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: LIVER, NASAL MUCOSA AND EPIDIDYMIS.
 CC -!- DEVELOPMENTAL STAGE: LIVER FROM ADULT FEMALE RATS CONTAINS ABOUT
 CC 10-FOLD GREATER LEVELS OF YC2 THAN IS FOUND IN LIVER FROM ADULT
 CC MALE RATS.
 CC -!- INDUCTION: BY ETHOXYQUIN, OLTIPIRAZ, BUTYLATED HYDROXYANISOLE, AND
 CC PHENOBARBITOL.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X78847; CAA55404.1; -;
 DR EMBL; S72506; -; NOT_ANNOTATED_CDS.
 DR EMBL; S82820; AAB46796.1; -;
 DR HSSP; P08263; IGSD.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR004045; GST_Nterm.
 DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF02798; GST_N; 1.
 KW Transferase; Multigene family.
 FT INIT_MET 0
 FT CONFLICT 101 101 L -> I (IN REF. 2).
 FT CONFLICT 183 183 L -> K (IN REF. 2).
 SQ SEQUENCE 220 AA; 25216 MW; DE8D9D43365664F5 CRC64;
 Query Match 0.6%; Score 7; DB 1; Length 220;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1089 KDKARNR 1095
 Db 124 KDKARNR 130
 RESULT 28
 GTC2_RAT
 ID GTC2_RAT STANDARD; PRT; 220 AA.
 AC P46418;
 DT 01-NOV-1995 (Rel. 32, Created)

```
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
DB 124 KDKARNR 130
|||||

RESULT 29
KPYK_THELI
ID KPYK_THELI STANDARD; PRT; 220 AA.
AC Q56301;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate kinase (EC 2.7.1.40) (PK) (fragment).
GN PKI.
OS Thermococcus litoralis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=2265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 5473;
RX MEDLINE=95370164; PubMed=7642512;
RA Kletzin A., Mukund S., Kelley-Crouse T.L., Chan M.K., Rees D.C.,
RA "Molecular characterization of the genes encoding the tungsten-
RT containing aldehyde ferredoxin oxidoreductase from Pyrococcus
RT furiosus and formaldehyde ferredoxin oxidoreductase from Thermococcus
RT litoralis."
RL J. Bacteriol. 177:4817-4819(1995).
CC -!- CATALYTIC ACTIVITY: ATP + pyruvate -> ADP + phosphoenolpyruvate.
CC -!- PATHWAY: GLYCOLYSIS; final step.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PYRUVATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X83963; CAA58793.1; -
CC HSP; P11974; LPKN.
CC InterPro: IPR001697; Pyruvate_kinase.
CC Pfam: PF00224; PK; 1.
CC ProDom: PD001009; Pyruvate_kinase; 1.
CC PROSITE: PS00110; PYRUVATE_KINASE; PARTIAL.
KW Transferase; Kinase; Glycolysis; Magnesium.
FT NON_TER 220 220
SQ SEQUENCE 220 AA; 24338 MW; A60B906BC742FCFAA CRC64;

Query Match 0.6%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 VKSFLEK 25
DB 209 VKSFLEK 215
|||||

RESULT 30
EUTQ_SALTY
ID EUTQ_SALTY STANDARD; PRT; 229 AA.
AC Q9ZEV5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ethanolamine utilization protein eutQ.
GN EUTQ OR STM2468.

Salmonella typhimurium.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=99395039; PubMed=10464203;
RA Kofoid E.C., Rappleye C.A., Stojilkovic I., Roth J.R.;
RA "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium
RT encodes five homologues of carboxysome shell proteins."
RL J. Bacteriol. 181:5317-5329(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -!- PATHWAY: Ethanolamine utilization.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF093749; AAC78113.1; -
CC EMBL; AE008811; AAL21362.1; -
CC Sygene; SG10633; eutQ.
KW Complete proteome.
SQ SEQUENCE 229 AA; 24992 MW; 75ADAC10E5F88279 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 TANDIRA 473
DB 6 TANDIRA 12
|||||

RESULT 31
RADCBACSU
ID RADCBACSU STANDARD; PRT; 231 AA.
AC Q02170;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA repair protein radC homolog.
KW RAD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93015731; PubMed=1400224;
RA Levin P.A., Margolis P.S., Setlow P., Losick R., Sun D.;
RA "Identification of Bacillus subtilis genes for septum placement and
RT shape determination."
RL J. Bacteriol. 174:6717-6728(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93259962; PubMed=8387996;
RA Butler Y.X., Abhayawardhane Y., Stewart G.C.;
```


Amplification of the Bacillus subtilis maf gene results in arrested septum formation.
J. Bacteriol. 175:3139-3145(1993).
[3]
SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Arvedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita Y., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Giuseppe G., Guy B.J., Haga K., Haeleth J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klier-Bianchard M., Klein C., Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rev M., Reynolds S., Rieger M., Rivoita C., Roche B., Rose M., Sadaie I., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis";
RL Nature 390:249-256(1997).
CC -1- FUNCTION: INVOLVED IN DNA REPAIR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RADC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC
CC EMBL; M96343; AAA22396.1; -
DR EMBL; L08793; AAA22583.1; -
DR EMBL; Z99118; CAB14764.1; -
DR PIR; B45239; B45239.
DR Subtilist; B010325; radc.
DR InterPro; IPR003583; HHH_1.
DR InterPro; IPR000445; HHH.
DR InterPro; IPR001405; RadC.
DR Pfam; PF00633; HHH; 1.
DR ProDom; PD007415; RadC; 1.
DR SMART; SMO0278; HHH; 1.
DR TIGRFAMs; TIGR00608; radc; 1.
DR PROSITE; PS01302; RADC; 1.
KW DNA repair; Complete proteome.
SQ SEQUENCE 231 AA; 26146 MW; B93FC0528C5B49F4 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 VIGDKKF 442
| | | | |
DB 216 VIGDKKF 222

RESULT 32

1431_ENTHI
ID 1431_ENTHI STANDARD; PRT; 239 AA.
AC P42848;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 14-3-3 protein 1 (14-3-3-1).
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=HM-1;IMSS;
RA Samuelson J., Shen P., Meckler G., Descoteaux S., Fu H., Lohia A.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE 14-3-3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC
CC EMBL; U13418; AAA80185.1; -
DR EMBL; P29312; 1A38.
DR InterPro; IPR000308; 14-3-3.
DR Pfam; PF00244; 14-3-3; 1.
DR PRINTS; PR00305; 1433ZETA.
DR ProDom; PD000600; 14-3-3; 1.
DR SMART; SMO0103; 14_3; 1.
DR PROSITE; PS00796; 1433_1; 1.
DR PROSITE; PS00797; 1433_2; 1.
KW Multigene family.
SQ SEQUENCE 239 AA; 27338 MW; 6BDE1496C8428FFC CRC64;

Query Match 0.6%; Score 7; DB 1; Length 239;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 409 LFPKAST 415
| | | | |
DB 108 LFPKAST 114

RESULT 33

CCS_MOUSE
ID CCS_MOUSE STANDARD; PRT; 274 AA.
AC Q9WU84; Q9CRJ9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper chaperone for superoxide dismutase (Superoxide dismutase copper chaperone).
DE CCS OR CCSO.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Bartnikas T.B., Waggoner D.J., Gitlin J.D.;
RT "The mouse copper chaperone for superoxide dismutase (CCS).";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Brain;
RX MEDLINE=20237523; PubMed=10773661;
RA Moore S.D., Chen M.M., Cox D.W.;

RT *Cloning and mapping of murine superoxide dismutase copper chaperone
 RT (ccsd) and mapping of the human ortholog.;"
 RL Cytogenet. Cell Genet. 88:35-37(2000).
 RN [3]

RP SEQUENCE OF 103-274 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Gishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;

RT *Functional annotation of a full-length mouse cDNA collection.;"
 RL Nature 409:685-690(2001).

CC -1- FUNCTION: Delivers copper to copper zinc superoxide dismutase
 CC (SOD1).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- TISSUE SPECIFICITY: Ubiquitous.

CC -1- SIMILARITY: CONTAINS 1 HMA DOMAIN.

CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CU-ZN
 CC SUPEROXIDE DISMUTASE FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

DR EMBL: AF121906; AAD23832.1; -
 DR EMBL: AF173379; AAF70342.1; -
 DR EMBL: AK010264; BAB26806.1; -
 DR HSSP: P15107; IXS0.

DR MGD: MGI:1333783; Ccsd.

DR InterPro: IPR001934; HeavyMe.transpt.

DR InterPro: IPR001424; SOD_CU_ZN.

DR Pfam: PF00080; sodecu; 1.

DR Pfam: PF00403; HMA; 1.

DR PRINTS: PR00068; CUZNDISMUTASE.

DR PRODOM: PD000469; SOD_CU_ZN; 1.

DR PROSITE: PS01047; HMA_1; FALSE_NEG.

DR PROSITE: PS0846; HMA_2; 1.

DR PROSITE: PS00087; SOD_CU_ZN_1; FALSE_NEG.

DR PROSITE: PS00332; SOD_CU_ZN_2; 1.

KW Chaperone; Copper; Zinc; Metal-binding.

FT DOMAIN 12 75

FT HMA.

FT DOMAIN 88 234 SUPEROXIDE DISMUTASE-LIKE.

FT METAL 22 22 COPPER (POTENTIAL).

FT METAL 25 25 COPPER (POTENTIAL).

FT METAL 130 130 COPPER (BY SIMILARITY).

FT METAL 132 132 COPPER (BY SIMILARITY).

FT METAL 147 147 COPPER AND ZINC (BY SIMILARITY).

FT METAL 155 155 ZINC (BY SIMILARITY).

FT METAL 164 164 ZINC (BY SIMILARITY).

FT METAL 167 167 ZINC (BY SIMILARITY).

FT METAL 201 201 COPPER (BY SIMILARITY).

SQ SEQUENCE 274 AA; 28911 MW; 19DCE48376C9D5A2 CRC64;

Query Match

0.6%; Score 7; DB 1; Length 274;

Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 GTVCALE 36
 |||||
 Db 9 GTVCALE 15

RESULT 34

UPK_BACSU

ID UPK_BACSU STANDARD; PRT; 276 AA.

AC P94507;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative undecaprenol kinase (EC 2.7.1.66) (Bacitracin resistance

DE protein).

DE UPK OR BACA.

GN Bacillus subtilis.

OS Bacillales; Bacillaceae; Bacillus.

OC Bacillales; Firmicutes;

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,

RA Guseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parto V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Roche E., Roche M., Sadaie Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Viari A., Wiputt R., Wedler E., Wedler H., Weitzenecker T.,

RA Winters P., Wiputt A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,

RT "The complete genome sequence of the Gram-positive bacterium Bacillus

RT subtilis.;"

RL Nature 390:249-256(1997).

RN [2]

RP SEQUENCE OF 31-276 FROM N.A.

RC STRAIN=PB1831;

RA De Rossi E.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Probably phosphorylates undecaprenol to undecaprenyl

CC phosphate. Confers resistance to bacitracin (By similarity).

CC -1- CATALYTIC ACTIVITY: ATP + undecaprenol = ADP + undecaprenyl

CC phosphate.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -1- MISCELLANEOUS: Bacitracin is thought to be involved in inhibition

CC of peptidoglycan synthesis by sequestering undecaprenyl

CC diphosphate reducing the pool of lipid carrier available.

CC -1- SIMILARITY: BELONGS TO THE UPK FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z91119; CAB15093.1; -
DR EMBL; U87792; AAB47703.1; ALT_INIT.
DR Subtilisin; BG13951; upk.
DR InterPro; IPR003824; BACA.
DR Pfam; PF02673; BACA; 1.
DR TIGRFAMs; TIGR00753; undec_kin_baca; 1.
KW Transferase; Kinase; Antibiotic resistance; Transmembrane;
FT TRANSMEM 49 58 POTENTIAL.
FT TRANSMEM 89 111 POTENTIAL.
FT TRANSMEM 117 139 POTENTIAL.
FT TRANSMEM 194 216 POTENTIAL.
FT TRANSMEM 226 248 POTENTIAL.
FT TRANSMEM 255 274 POTENTIAL.
SQ SEQUENCE 276 AA; 30345 MW; 2ED137A85F5833BD CRC64;

Query Match 0.6%; Score 7; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 966 FEDYIDE 972
|||||
DB 111 FEDYIDE 117

RESULT 35
Y620_METJA STANDARD; PRT; 291 AA.
ID Y620_METJA
AC Q58037;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0620.
GN MJ0620.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
SQ SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2651 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8698087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO M.JANNASCHII MJ1001 AND TO E.COLI RIBOSOMAL PROTEIN
CC S6 MODIFICATION PROTEIN (RIMK).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U67510; AAB98615.1; -
DR TIGR; MJ0620; -
DR InterPro; IPR004666; RimK_fam.

DR TIGRFAMs; TIGR00768; rimk_fam; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 291 AA; 33277 MW; 2B6A343829A0A3B0E CRC64;

Query Match 0.6%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 PSWIGLS 210
|||||
DB 264 PSWIGLS 270

RESULT 36
UL07_HSVEB STANDARD; PRT; 303 AA.
ID UL07_HSVEB
AC P28945;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Gene 55 protein.
GN 55.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL7,
CC EHV-1 55, VZV 53, EBV BBRF2, HCMV UL103 AND HVS-1 42.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M86664; AAB02490.1; -
DR PIR; A36801; WZBEE7.
DR InterPro; IPR002600; Herpes_UL7.
DR Pfam; PF01677; Herpes_UL7; 1.
SQ SEQUENCE 303 AA; 33854 MW; 4449A86C70ACA0ED CRC64;

Query Match 0.6%; Score 7; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1107 LSHRLVL 1113
|||||
DB 130 LSHRLVL 136

RESULT 37
GLSA_STRCO STANDARD; PRT; 307 AA.
ID GLSA_STRCO
AC P57755;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable glutaminase SCO7049 (EC 3.5.1.2).
GN SCO7049 OR SC461.15.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- CATALYTIC ACTIVITY: L-glutamine + H(2)O -> L-glutamate + NH(3).
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL391039; CAC01547.1; .
KW Hydrolase; Complete proteome.
SQ SEQUENCE 307 AA; 32555 MW; CCF51DFAE88D2079 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 307;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 533 GDFARV 539
DB 247 GDFARV 253
|||||

RESULT 38
LIP_PSESP
ID LIP_PSESP STANDARD; PRT; 311 AA.
AC P26877;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lactonizing lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).
GN LIP.
OS Pseudomonas sp. (strain 109).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-45.
RX MEDLINE=92011544; PubMed=1917947;
RA Ihara F., Kageyama Y., Hirata M., Nihira T., Yamada Y.;
RT "Purification, characterization, and molecular cloning of lactonizing
RT lipase from Pseudomonas species.";
RL J. Biol. Chem. 266:18135-18140(1991).
CC -!- FUNCTION: CATALYZES THE SYNTHESIS OF MACROCYCLIC LACTONES IN
CC ANHYDROUS ORGANIC SOLVENTS.
CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O -> diacylglycerol + a
CC fatty acid anion.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS
CC LIPASE FAMILY.
CC -!- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC

CC EMBL; D10166; BAA01035.1; .
DR PIR; A40943; A40943.
DR HSP; P22088; 3LIP.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR000734; Lipase.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00581; abhydrolase; 1.
DR PROSITE; PS00120; LIPASE_SER; 1.
KW Hydrolase; Lipid degradation; Signal.
FT SIGNAL 1 26
FT CHAIN 27 311 LACTONIZING LIPASE.
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 235 235 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 277 277 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 311 AA; 32737 MW; 27AC2F3DD3B334D1 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 311;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 FLSSGSA 139
DB 170 FLSSGSA 176
|||||

RESULT 39
ISPH_BACHD
ID ISPH_BACHD STANDARD; PRT; 314 AA.
AC Q9KD37;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ISPH protein.
GN ISPH OR LYTB OR BH1382.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
CC (DMAPP) (By similarity).
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
CC (last) step.
CC -!- SIMILARITY: BELONGS TO THE ISPH FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001511; BAB05101.1; .
DR InterPro; IPR003453; LytB.
DR Pfam; PF02401; LytB; 1.
DR TIGRFAMs; TIGR00216; lytB; 1.
KW Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 314 AA; 34841 MW; BC0C12493B22698F CRC64;

Query Match 0.6%; Score 7; DB 1; Length 314;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 56 ISADKII 62
Db 159 ISADKII 165
|||||

RESULT 40
VG2_SPV4
ID VG2_SPV4 STANDARD; PRT; 320 AA.
AC P11334;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last annotation update)
DE Gene 2 protein.
GN 2.
OS Spiroplasma virus 4 (SPV4).
OC Viruses; ssDNA viruses; Microviridae; Spiromicrovirus.
OX NCBI_TaxID=10855;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86032809; PubMed=2822658;
RA Renaudin J., Pascarel M.-C., Bove J.-M.;
RT "Spiroplasma virus 4: nucleotide sequence of the viral DNA, regulatory signals, and proposed genome organization.";
RL J. Bacteriol. 169:4950-4961(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17988; -; NOT_ANNOTATED_CDS.
DR PIR; H29825; G2BPSV.
SQ SEQUENCE 320 AA; 38090 MW; B4C1B2BCACIA2058 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 849 NYIVNDS 855
Db 36 NYIVNDS 42
|||||

RESULT 41
Y526_CHLPN
ID Y526_CHLPN STANDARD; PRT; 329 AA.
AC Q92826; Q9JQ87;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CP0526/CP0226/CPJ0526.
GN CPNG526 OR CP0226 OR CPJ0526
OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206506; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W., Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,

```

```

RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- SIMILARITY: BELONGS TO THE SIS FAMILY. GUTQ/KPSF SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001637; AAD18666.1; -
DR EMBL; AE002183; AAF38092.1; -
DR EMBL; AF002547; BAA98732.1; -
DR TIGR; CP0226; -
DR InterPro: IPR000644; CBS_domain.
DR InterPro: IPR004800; KpsF.
DR InterPro: IPR001347; SIS.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF01380; SIS; 1.
DR SMART; SM00116; CBS; 1.
DR TIGRFAMs; TIGR00393; kpsF; 1.
KW Hypothetical protein; ATP-binding; Repeat; CBS domain;
FT NP_BIND 52 57 ATP (POTENTIAL).
FT DOMAIN 204 258 CBS 1.
FT DOMAIN 272 327 CBS 2.
FT SEQUENCE 329 AA; 35650 MW; 27A3023B6ABE6D2F CRC64;

Query Match 0.6%; Score 7; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 KQKEAYD 629
Db 17 KQKEAYD 23
|||||

RESULT 42
PDXA_FUSNN
ID PDXA_FUSNN STANDARD; PRT; 332 AA.
AC Q8RGU0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 4-hydroxythreonine-4-phosphate dehydrogenase (EC 1.1.1.262) (4-phosphohydroxy)-L-threonine dehydrogenase).
GN PDXA OR FN0226.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,

```

```

RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fontein M., Kyprides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-
CC (phosphohydroxy)-L-threonine (HTP) into 2-amino-3-oxo-4-
CC form 1-amino-3-(phosphohydroxy)propan-2-one (3-amino-2-oxopropyl
CC phosphate) (by similarity).
CC -!- CATALYTIC ACTIVITY: 4-(phosphonoxy)-threonine + NAD(+) = 2-
CC amino-3-oxo-4-phosphonoxybutyrate + NADH
CC -!- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and
CC pyridoxal phosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE PDXA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE010535; AAL94432.1; -
CC Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
KW Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 332 AA; 36347 MW; A5032E3F6E7081FC CRC64;
-----
Query Match 0.6%; Score 7; DB 1; Length 332;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 711 TLEYGVV 717
| | | | |
DB 85 TLEYGVV 91

RESULT 43
RPOA_PSEAE
ID RPOA_PSEAE STANDARD; PRT; 333 AA.
AC 052760;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (Transcriptase
DE alpha chain) (RNA polymerase alpha subunit).
GN RPOA OR PA4238.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FR01;
RX MEDLINE=99296583; PubMed=10368149;
RA Ma J.-F., Ochsner U.A., Klotz M.G., Nanayakkara V.K., Howell M.L.,
RA Johnson Z., Posey J.E., Vasili M.L., Moraco J.J., Hassett D.J.;
RT "Bacterioferritin A modulates catalase A (KatA) activity and
RT resistance to hydrogen peroxide in Pseudomonas aeruginosa."
RL J. Bacteriol. 181:3730-3742(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."

```

```

RL Nature 406:959-964(2000).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: CONSISTS OF A SIGMA FACTOR AND THE RNAP CORE ENZYME WHICH
CC IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, 1 BETA' CHAIN AND 1
CC OMEGA CHAIN (BY SIMILARITY).
CC -!- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF
CC CORE RNAP, WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH
CC TRANSCRIPTIONAL REGULATORS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF047025; AAC03116.1; -
CC EMBL; AE004841; AAG07626.1; -
CC HSP; P00574; 1COO
CC InterPro: IPR001700; RNA_pol_A_bac.
CC Pfam: PF01000; RNA_pol_A_bac; 1.
CC Pfam: PF03118; RNA_pol_A_CTD; 1.
CC ProDom: PD001179; RNA_pol_A_bac; 1.
CC Transferrase; Transcription; DNA-directed RNA polymerase;
KW Complete proteome.
FT CONFLICT 326 333 KKDKATA -> TERROGHCLIVVTERKVKWGIEPCAIVK
FT VVTV (IN REF. 1).
SQ SEQUENCE 333 AA; 36649 MW; 50706D2926207CA9 CRC64;
-----
Query Match 0.6%; Score 7; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 832 IQLDHDV 838
| | | | |
DB 115 IQLDHDV 121

RESULT 44
Y094_CAEEL
ID Y094_CAEEL STANDARD; PRT; 335 AA.
AC P41844;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 40.1 kDa protein T20B12.4 in chromosome III.
GN T20B12.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: STRONG, TO C.ELEGANS K06H7.2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U10401; AAA19057.1; -

```

DR WormPep: T20B12.4; CE01410.

KW Hypothetical protein.

SQ SEQUENCE 335 AA; 40139 MW; 9DE51B219062E8E7 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 335;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 DFSQIKN 483

Db 183 DFSQIKN 189

RESULT 45

RUVB_RICCN STANDARD; PRT; 342 AA.

AC Q92187;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Holliday junction DNA helicase ruvb.

GN RUVB OR RC0533.

OS Rickettsia conorii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_taxid=781;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Malish 7;

RX MEDLINE=21442074; PubMed=11557893;

RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,

RA Samson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,

RA Raoult D.;

PT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";

RL Science 293:2093-2098(2001).

CC -!- FUNCTION: The ruva-ruvb complex in the presence of ATP renatures

CC cruciform structure in supercoiled DNA with palindromic sequence,

CC indicating that it may promote strand exchange reactions in

CC homologous recombination. RuvAB is an helicase that mediates the

CC Holliday junction migration by localized denaturation and

CC reannealing (By similarity).

CC -!- SUBUNIT: Forms a complex with ruva (By similarity).

CC -!- SIMILARITY: BELONGS TO THE RUVB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AE008615; AAL03071.1; ALT_INIT.

CC InterPro: IPR003593; AAA_ATPase.

CC InterPro: IPR003959; AAA_ATPase_cent.

CC InterPro: IPR004605; RuvB.

CC Pfam: PF00004; AAA; 1.

CC SMART: SM00382; AAA; 1.

CC TIGRfam: TIGR00635; ruvb; 1.

CC DNA repair; SOS response; ATP-binding; DNA recombination; Helicase;

CC Complete proteome.

FT NP_BIND 57

SQ SEQUENCE 342 AA; 38309 MW; 83891BDB0C829C28 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 342;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 681 RASKLLD 687

Db 190 RASKLLD 196

RESULT 46

RUVB_RICPR

ID RUVB_RICPR STANDARD; PRT; 342 AA.

AC Q92DE5;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Holliday junction DNA helicase ruvb.

GN RUVB OR RP386.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_taxid=782;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Madrid E;

RX MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,

RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of

RT mitochondria";

RL Nature 396:133-140(1998).

CC -!- FUNCTION: THE RUVB-RUVB COMPLEX IN THE PRESENCE OF ATP RENATURES

CC CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,

CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN

CC HOMOLOGOUS RECOMBINATION. RUVAB IS AN HELICASE THAT MEDIATES THE

CC HOLLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND

CC REANNEALING (BY SIMILARITY).

CC -!- FUNCTION: PARTICIPATES TO UV-TOLERANCE OF SYNECHOCYSTIS PCC 6803.

CC -!- SUBUNIT: FORMS A COMPLEX WITH RUVB.

CC -!- SIMILARITY: BELONGS TO THE RUVB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AJ235271; CAA14843.1; -

CC InterPro: IPR003593; AAA_ATPase.

CC InterPro: IPR003959; AAA_ATPase_cent.

CC Pfam: PF00004; AAA; 1.

CC SMART: SM00382; AAA; 1.

CC TIGRfam: TIGR00635; ruvb; 1.

CC DNA repair; SOS response; ATP-binding; DNA recombination; Helicase;

CC Complete proteome.

FT NP_BIND 57

SQ SEQUENCE 342 AA; 38435 MW; D2F9290E84079BA5 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 342;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 681 RASKLLD 687

Db 190 RASKLLD 196

RESULT 47

NOOB_PARDE

ID NOOB_PARDE STANDARD; PRT; 345 AA.

AC P29920;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE NADH-ubiquinone oxidoreductase chain 8 (EC 1.6.5.3) (NADH

DE dehydrogenase 1, chain 8) (NDH-1, chain 8).

GN NOOB.

OS Paracoccus denitrificans.

```

CC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
CC Paracoccus.
OX NCBI_TaxID=266;
RN [1]
RP SEQUENCE FROM N.54;
RX STRAIN=ATCC 13543;
RA MEDLINE=93136200; PubMed=8422400;
RA Xu X., Matsuno-Yagi A., Yagi T.;
RT "DNA sequencing of the seven remaining structural genes of the gene
RT cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT of Paracoccus denitrificans.";
RL Biochemistry 32:968-981(1993).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS, SUBUNITS N007-14
CC CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L02354; AAA25592.1; -
CC PIR; C45456; C45456
CC InterPro; IPR001694; Resp_NADH_dhl.
CC Pfam; PF00146; NADHdh; 1.
CC PROSITE; PS00667; COMPLEX1_ND1_1; 1.
CC PROSITE; PS00668; COMPLEX1_ND1_2; 1.
CC Oxidoreductase; NAD; Ubiquinone; Transmembrane.
CC TRANSMEM 15 35 POTENTIAL.
CC TRANSMEM 82 102 POTENTIAL.
CC TRANSMEM 115 135 POTENTIAL.
CC TRANSMEM 161 181 POTENTIAL.
CC TRANSMEM 190 210 POTENTIAL.
CC TRANSMEM 253 273 POTENTIAL.
CC TRANSMEM 278 298 POTENTIAL.
CC TRANSMEM 309 329 POTENTIAL.
CC SEQUENCE 345 AA; 38751 MW; E33B667E569506B4 CRC64;
SQ
Query Match 0.6%; Score 7; DB 1; Length 345;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 671 PFLSMML 677
DB 88 PFLSMML 94
[1]
RESULT 48
YJ9L YEAST
ID YJ9L YEAST STANDARD; PRT; 358 AA.
AC P47175;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 41.2 kDa protein in RPS4A-BAR2 intergenic region.
GN YJ9L147W OR J7204.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Scarcez T.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z49646; CAA89680.1; -
CC HSSP; P22121; ZHTS.
CC TRANSFAC; T034449; -
CC SGD; S0003908; YJ9L147W.
CC InterPro; IPR002341; HSF_DNA_bind.
CC InterPro; IPR002341; HSF_ETs.
CC Pfam; PF00447; HSF_DNA-bind; 1.
CC ProDom; PD001788; HSF_DNA_bind; 1.
CC SMART; SM00415; HSF; 1.
CC PROSITE; PS00434; HSF_DOMAIN; FALSE_NEG.
CC Hypothetical protein; Nuclear protein; DNA-binding.
FT DNA_BIND 12 116 BY SIMILARITY.
SQ SEQUENCE 358 AA; 41192 MW; E2964157FC952D42 CRC64;
Query Match 0.6%; Score 7; DB 1; Length 358;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 605 QLITLLS 611
DB 347 QLITLLS 353
[1]
RESULT 49
BIOB_SYNY3
ID BIOB_SYNY3 STANDARD; PRT; 362 AA.
AC P73538;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).
GN BIOB OR SLR1364.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Watsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Tamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC -1- PATHWAY: Biotin biosynthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90907; BAA17578.1; -
CC InterPro; IPR002684; Biotin_synth.
CC Pfam; PF01792; Biotin_synth; 1.
CC TIGPfam; TIGR00433; bioB; 1.
CC Biotin biosynthesis; Iron-sulfur; Transferrase; Complete proteome.
KW METAL 88
FT

```


FT METAL 92 92 IRON-SULFUR (POTENTIAL).
FT METAL 95 95 IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 362 AA; 33349 MW; BF7A16F81CB68B4 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 994 KTEAEIL 1000
Db 112 KTEAEIL 118

RESULT 50

YPC4_CABEL STANDARD; PRT; 371 AA.
AC Q11181;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 41.9 kDa protein C05D10.4 in chromosome III.
GN C05D10.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U13645; AAA20988.1; -.
DR WormPep; C05D10.4; CE01127.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00536; SAM; 1.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS0105; SAM_DOMAIN; 1.
KW Hypothetical protein.
FT DOMAIN 223 286
SQ SEQUENCE 371 AA; 41906 MW; 6CE79B87D8DD390C CRC64;

Query Match 0.6%; Score 7; DB 1; Length 371;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GSTEVS 121
Db 113 GSTEVS 119

Search completed: November 6, 2002, 03:35:05
Job time : 100 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 16:47:42 ; Search time 702 Seconds
(without alignments)
11968.954 Million cell updates/sec

Title: US-09-782-874-1
Perfect score: 3731
Sequence: 1 GAATATCTTACTACTT.....AGTTTCATCTTCTCTCTAAA 3731

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3731	100.0	3731	22 AAD04370	Tomato RNA-directed
2	3731	100.0	3731	23 AAS17837	Tomato RNA-directe
3	3721.4	99.7	3731	23 AAS17845	Tomato RNA-directe
4	1396	37.4	3807	21 AAC63742	Soybean RNA-direct
5	987.8	26.5	3737	21 AAC63739	Maize RNA-directed
6	345	9.2	3901	21 AAC63740	Maize RNA-directed
7	305.8	8.2	3591	22 AAH77704	Nucleotide sequenc
8	304.2	8.2	2816	21 AAC63741	Rice RNA-directed
9	282	7.6	1281	21 AAC63738	Maize RNA-directed

10	276.2	7.4	740	21	AAC63743	Maize RNA-directed
11	220.4	5.9	6863	22	AAH77703	Nucleotide sequenc
12	106.8	2.9	282	24	ABL71773	Corn tassell-derive
13	95.4	2.6	274	24	ABL71722	Corn tassell-derive
14	85.8	2.3	349	21	AAC63746	Soybean RNA-direct
15	71.4	1.9	510	21	AAC63745	Soybean RNA-direct
16	57.2	1.5	906	17	AAR18078	Potato calmodulin
17	49	1.3	8045	21	AAAC65171	Neurospora crassa
18	41.2	1.1	339	21	AAC18542	Human secreted pro
19	41.2	1.1	7156	24	ABL59541	Human interleukin
20	40.4	1.1	23071	22	AAS26699	Human genomic DNA
21	40.2	1.1	1904	22	AAK05148	Human brain expres
22	40.2	1.1	14041	22	AAR48024	Internal control B
23	40	1.1	4590	22	AAR48065	yeast AOD9604-asso
24	40	1.1	162450	21	AAR86967	Retinoblastoma bin
25	39.4	1.1	10716	24	ABL33418	Human immune syste
26	39.2	1.1	2073	23	AAS54725	Staphylococcus aur
27	39.2	1.1	2109	23	AAS51973	Staphylococcus aur
28	39.2	1.1	6012	24	ABL70327	Chemically treated
29	39.2	1.1	6012	24	AAS61274	Human gene regulat
30	39.2	1.1	6012	24	ABK31370	Signal transductio
31	39	1.0	7580	24	ABL33223	Human immune syste
32	38.8	1.0	11790	24	ABL32542	Human immune syste
33	38.6	1.0	2570	23	ABL22608	Drosophila melanog
34	38.6	1.0	4803	23	ABL16961	Drosophila melanog
35	38.6	1.0	6079	24	ABL32421	Human immune syste
36	38.6	1.0	6225	23	ABL20912	Drosophila melanog
37	38.6	1.0	7027	23	ABL16960	Drosophila melanog
38	38.4	1.0	5179	24	ABL33996	Human immune syste
39	38.4	1.0	5179	24	ABK28403	DNA transcription
40	38.2	1.0	1537	21	AAC39240	Arabidopsis thalia
41	38.2	1.0	15667	24	ABL34146	Human immune syste
42	38.2	1.0	113515	24	ABL34174	Human immune syste
43	38	1.0	540	24	ABN64825	Human cancer relat
44	38	1.0	4236	24	AAI67659	Nucleotide sequenc
45	38	1.0	5191	21	AAF20989	Human low adenosin

ALIGNMENTS

RESULT 1
AAD04370

ID AAD04370 standard; cdna; 3731 BP.

AC AAD04370;

DT 04-JUL-2001 (first entry)

XX Tomato RNA-directed RNA polymerase (RdRP) cdna.

DE Tomato: gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
transgenic plant; tissue culture; plant breeding; therapy; C-protein; ss.

XX Lycopersicon esculentum.

XX Key Location/Qualifiers

FT CDS 194..3538

FT /*tag= a

FT /product= "Tomato C-protein having RNA-directed RNA
polymerase (RdRP) activity"

XX US6218142-B1.

PN 17-APR-2001.

XX 05-MAR-1997; 97US-0811583.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.
PA (RIED/) RIEDEL L.

Wassenegger M, Riedel L, Schiebel W, Sanger HL;
WPI: 2001-289830/30.
P-PSDB; AAE00897.

New nucleic acid molecules encoding polypeptides with RNA-directed RNA polymerase enzymatic activity, useful in modulating gene expression in plants, humans and animals, as well as in plant cell/tissue cultures or plant breeding.

Claim 1; Column 25-34; 31pp; English.

The present sequence is a cDNA encoding tomato C-protein having RNA-directed RNA polymerase (RdRP) activity. This protein is capable of RNA-directed RNA synthesis, thus using RNA as a template for synthesizing complementary RNA molecules. RdRP nucleic acid is useful for modulating gene expression in plants, humans and animals. This may lead to various physiological, developmental and/or morphological changes. Transgenic plants containing RdRP nucleic acid is especially useful in plant cell or tissue cultures and in plant breeding. RdRP is useful in gene therapy, particularly for treating a disease that is caused by the undesirable expression or overexpression of a gene.

Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 other;

Query Match 100.0%; Score 3731; DB 22; Length 3731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GAAATATCTTACTTACTTACCAGGAGTGAATCATCTACTCCCTCAAGTCTTTGTGT 60
1 GAATATCTTACTTACTTACCAGGAGTGAATCATCTACTCCCTCAAGTCTTTGTGT 60

61 GTTGTGATAATAAATTGGTGTGCTTCAGTTTCCAGTCACTACTGCGGGTGTATTTAT 120
61 GTTGTGATAATAAATTGGTGTGCTTCAGTTTCCAGTCACTACTGCGGGTGTATTTAT 120

121 TTGTGATAAATCTGAGGAGTATCCAGTTGGTGTAGCATTTGAAGTCAAGTCACTT 180
121 TTGTGATAAATCTGAGGAGTATCCAGTTGGTGTAGCATTTGAAGTCAAGTCACTT 180

181 GGAATTTGGCTACATGGAAGACAAATTCAGGTTTTCGGATTCCTTATCTTCTCTGC 240
181 GGAATTTGGCTACATGGAAGACAAATTCAGGTTTTCGGATTCCTTATCTTCTCTGC 240

241 GGAAGTGTAAAGTCACTTTAGAGAAATATACAGGATATGGAATGATGTGATGGA 300
241 GGAAGTGTAAAGTCACTTTAGAGAAATATACAGGATATGGAATGATGTGATGGA 300

301 GGTAAACAGTCCAAAGGAGGATCTAGAGCAATTTGCCAAAGTTCAAATTTGCCGACACAT 360
301 GGTAAACAGTCCAAAGGAGGATCTAGAGCAATTTGCCAAAGTTCAAATTTGCCGACACAT 360

361 AAGTCTGACAAATCATCATCTTTGGCTAATAACAGGCTGATTTTGGCTCTTCTTATTT 420
361 AAGTCTGACAAATCATCATCTTTGGCTAATAACAGGCTGATTTTGGCTCTTCTTATTT 420

421 GAAGGCTTGGGAAATGAAATGATATGTCCAACTGCGGCGCATATGTGATCATAGATGGA 480
421 GAAGGCTTGGGAAATGAAATGATATGTCCAACTGCGGCGCATATGTGATCATAGATGGA 480

481 TGGCATACATTTGAATTTCCGATGTCAGATATCAGATGACAAAGTTTCCAGTTGGGAAG 540
481 TGGCATACATTTGAATTTCCGATGTCAGATATCAGATGACAAAGTTTCCAGTTGGGAAG 540

541 TACAGAAGTTTCAATTCATTTGGCATTTGGATTTGAAGAAATTTTCTTTTATCTAG 600
541 TACAGAAGTTTCAATTCATTTGGCATTTGGATTTGAAGAAATTTTCTTTTATCTAG 600

601 TGGTTTCACTGACTATAAATCTACGCTTTCATATGAAATATATGCGAGGTTGTGCTCCA 660
601 TGGTTTCACTGACTATAAATCTACGCTTTCATATGAAATATATGCGAGGTTGTGCTCCA 660

661 TCGTCCATATGTCACAAATGCTCAGTTTCTCCTCATACAGTATTTGTGTCTCTCGGAT 720
661 TCGTCCATATGTCACAAATGCTCAGTTTCTCCTCATACAGTATTTGTGTCTCTCGGAT 720

721 CATTAAAGAGACTTGAAGAACTCCTGTTTATAGCTTCTTTAAGNAACCTCGATGACAGT 780
721 CATTAAAGAGACTTGAAGAACTCCTGTTTATAGCTTCTTTAAGNAACCTCGATGACAGT 780

781 GGTGAGGACAAACAGATTTCCCTCCCATCTTGGATAGGCTATCTTCTAGCTATGTTTGA 840
781 GGTGAGGACAAACAGATTTCCCTCCCATCTTGGATAGGCTATCTTCTAGCTATGTTTGA 840

841 GTTCGTAGGAGTGTGCTTCTTCCAAATTCAGGAAAGTTTTTCCTACTATGCGAAGC 900
841 GTTCGTAGGAGTGTGCTTCTTCCAAATTCAGGAAAGTTTTTCCTACTATGCGAAGC 900

901 TGAAACAAATATTTACTTTACAGACGTTTTCACCTTTTTCGTTCTCTCAAAAATCGGCTCT 960
901 TGAAACAAATATTTACTTTACAGACGTTTTCACCTTTTTCGTTCTCTCAAAAATCGGCTCT 960

961 GGTTCCTCAATGTCACGCTCCCGAAGGATTTCAATTCCTCCTACAGATTTGTTTCAAAAT 1020
961 GGTTCCTCAATGTCACGCTCCCGAAGGATTTCAATTCCTCCTACAGATTTGTTTCAAAAT 1020

1021 TAGTTCTTTTGGTACAGCATGATGATACCTGGGCGAGCATTTAAATGTCTACTTTTCCG 1080
1021 TAGTTCTTTTGGTACAGCATGATGATACCTGGGCGAGCATTTAAATGTCTACTTTTCCG 1080

1081 ATTAGTTGATCTCGAAGGAGAAATGTGGCATGCAATGAGCATGCTTAGAGAAACTGTA 1140
1081 ATTAGTTGATCTCGAAGGAGAAATGTGGCATGCAATGAGCATGCTTAGAGAAACTGTA 1140

1141 CATTATAAAGGAGTGTGTTATGATCCCGTGGGTCACCTGAGCAGTATGATGGGTA 1200
1141 CATTATAAAGGAGTGTGTTATGATCCCGTGGGTCACCTGAGCAGTATGATGGGTA 1200

1201 TCTCAAGGTTAGACAACTCCAAATTCCTCGTCCATCACTTTAGATGATGGTGGTGA 1260
1201 TCTCAAGGTTAGACAACTCCAAATTCCTCGTCCATCACTTTAGATGATGGTGGTGA 1260

1261 TGTAAAGAGGCTTCTAGTAAACCATGCAAGTTTATTTTGGTCCAGAGGTTAATGT 1320
1261 TGTAAAGAGGCTTCTAGTAAACCATGCAAGTTTATTTTGGTCCAGAGGTTAATGT 1320

1321 TTCCAATCGGGTCTCCGCAATTTATCTGAAGACATAGATAAATTTCTTCTGTTTCTTT 1380
1321 TTCCAATCGGGTCTCCGCAATTTATCTGAAGACATAGATAAATTTCTTCTGTTTCTTT 1380

1381 TGTGATGAGGAGTGGAGAACTGTATTCTACAGACTTATTACCAAAAGCAAGTACTGG 1440
1381 TGTGATGAGGAGTGGAGAACTGTATTCTACAGACTTATTACCAAAAGCAAGTACTGG 1440

1441 AAGTGTGTGAGGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTGT 1500
1441 AAGTGTGTGAGGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTGT 1500

1501 AATTGTGATATAAATTTGAATTTCTTGCATTTTCATCGCCAGTTCGCGGATATTC 1560
1501 AATTGTGATATAAATTTGAATTTCTTGCATTTTCATCGCCAGTTCGCGGATATTC 1560

1561 AGTGTGATGTTTGGCATCAAGACTTGGCTTACTGCAAAATGATATAGAGCTTGGATGG 1620
1561 AGTGTGATGTTTGGCATCAAGACTTGGCTTACTGCAAAATGATATAGAGCTTGGATGG 1620

1621 TGATTTTCCAGATCAAGAAATGTCGCAAAATATGTCGCGAGACTTGGTCAATCTTTGG 1680
1621 TGATTTTCCAGATCAAGAAATGTCGCAAAATATGTCGCGAGACTTGGTCAATCTTTGG 1680

1681 TTCTCCAGAGACACTTGTAGTGTCTTGTAGGATGAGATTGAAGTTATTTCCCGATGTA 1740
1681 TTCTCCAGAGACACTTGTAGTGTCTTGTAGGATGAGATTGAAGTTATTTCCCGATGTA 1740

1741 GGTTATGGAACCAAGCTATGCTCTTTTCTGTAGGAATTTGGTAAATATCTGGTACTTGC 1800

```
Db 1741 GGTTCATGGAACCAAGTATGCTTTTCTGATGGAAATGGTAAATATCTGGTGACTTGG 1800
QY 1801 TCATAGAGTTGGCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGATTCGTTA 1860
Db 1801 TCATAGAGTTGGCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGATTCGTTA 1860
QY 1861 TGGTGATATAAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1920
Db 1861 TGGTGATATAAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1920
QY 1921 AAAGAGCATGTCGAATATGAATCAGACAAACATAAAGTTAGATGTCCTTGGATGGAGCAA 1980
Db 1921 AAAGAGCATGTCGAATATGAATCAGACAAACATAAAGTTAGATGTCCTTGGATGGAGCAA 1980
QY 1981 ATATCAGCGCTTGTATCTTAATCGTCAACTGATTAAGCTCTTGTCTACACTTGGAGTGA 2040
Db 1981 ATATCAGCGCTTGTATCTTAATCGTCAACTGATTAAGCTCTTGTCTACACTTGGAGTGA 2040
QY 2041 AGATGAAGTTCGGAACAGAGCAAAAGGAAGCTGTAGATCAGCTTGTATGCTATCTTGC 2100
Db 2041 AGATGAAGTTCGGAACAGAGCAAAAGGAAGCTGTAGATCAGCTTGTATGCTATCTTGC 2100
QY 2101 TGATCTTTGAAGGCACAGGAGGCTTTGGAATGTAGTCTCTCTGGAGAGAACACTAATAT 2160
Db 2101 TGATCTTTGAAGGCACAGGAGGCTTTGGAATGTAGTCTCTCTGGAGAGAACACTAATAT 2160
QY 2161 TCTCAAGCAATGCTAAACTGTGGTTATAAGCTGTATGCTGAGCCCTTCTTTCAATGAT 2220
Db 2161 TCTCAAGCAATGCTAAACTGTGGTTATAAGCTGTATGCTGAGCCCTTCTTTCAATGAT 2220
QY 2221 GTTGAACCTTCCGGCAGTCCAAAGTTGCTCGATTGGGAGCTAGATCAAGAATATTTAT 2280
Db 2221 GTTGAACCTTCCGGCAGTCCAAAGTTGCTCGATTGGGAGCTAGATCAAGAATATTTAT 2280
QY 2281 TCCAAATGAAGAACATGATGGATTTGGATGAATCCAGAACTTGGATATGGTCA 2340
Db 2281 TCCAAATGAAGAACATGATGGATTTGGATGAATCCAGAACTTGGATATGGTCA 2340
QY 2341 GGTGTTGTTCTAGTTTACTGTGCTGGACATGGAGAGTTTCTGACGATTTACATCCATT 2400
Db 2341 GGTGTTGTTCTAGTTTACTGTGCTGGACATGGAGAGTTTCTGACGATTTACATCCATT 2400
QY 2401 TAATACAGCAGATCCACCAACAGTAATTTCTGAGGGAATGTTGTTGGCAAA 2460
Db 2401 TAATACAGCAGATCCACCAACAGTAATTTCTGAGGGAATGTTGTTGGCAAA 2460
QY 2461 AAATCCATGCTTGCATCCTGGTGATATTCGTGTTTAAAGGCTGTAATGTTTCGAGCGCT 2520
Db 2461 AAATCCATGCTTGCATCCTGGTGATATTCGTGTTTAAAGGCTGTAATGTTTCGAGCGCT 2520
QY 2521 GCACCACTGGTAGATTGTTGTTATTCCTCAGAAAGGAAAAGACCTCATCCGAATGA 2580
Db 2521 GCACCACTGGTAGATTGTTGTTATTCCTCAGAAAGGAAAAGACCTCATCCGAATGA 2580
QY 2581 ATGTTCTGGAGGATTGATGGGATATCTACTTGTGTTGCTGGATCAAGACATGAT 2640
Db 2581 ATGTTCTGGAGGATTGATGGGATATCTACTTGTGTTGCTGGATCAAGACATGAT 2640
QY 2641 CCGGCCAAGGCAAGTCCAGCCGATGGAATATCTCCAGCACCAGCATACAGTTGGACCA 2700
Db 2641 CCGGCCAAGGCAAGTCCAGCCGATGGAATATCTCCAGCACCAGCATACAGTTGGACCA 2700
QY 2701 TGATGTCACAAATGAGGAAGTTCAGAGTACTTCCCAACTATATTGTGAATGACAGTTT 2760
Db 2701 TGATGTCACAAATGAGGAAGTTCAGAGTACTTCCCAACTATATTGTGAATGACAGTTT 2760
QY 2761 GGGAAATCATAGCAAAATGCCCATGCTGCTATTGTGAGAGAGAACCTGATATGGCCATGAG 2820
Db 2761 GGGAAATCATAGCAAAATGCCCATGCTGCTATTGTGAGAGAGAACCTGATATGGCCATGAG 2820
QY 2821 TGATCCATGCAAAAACTTCTGCTGAGCTCTTTTCAATTGAGTGGACTTTCCAAAGACTGG 2880
Db 2821 TGATCCATGCAAAAACTTCTGCTGAGCTCTTTTCAATTGAGTGGACTTTCCAAAGACTGG 2880
```

```
Db 2821 TGATCCATGCAAAAACTTCTGCTGAGCTCTTTTCAATTGAGTGGACTTTCCAAAGACTGG 2880
QY 2881 TGTTCGCCGTGAATATACCATCTCAGTTGCGCCCTTAAGATATACCAGACTTTCATGGATA 2940
Db 2881 TGTTCGCCGTGAATATACCATCTCAGTTGCGCCCTTAAGATATACCAGACTTTCATGGATA 2940
QY 2941 GCCGGCAAGCAGCAGCTATATCTCAGAAAGAGTTATTGAAAGCTTTTCAGGAAAGTGAA 3000
Db 2941 GCCGGCAAGCAGCAGCTATATCTCAGAAAGAGTTATTGAAAGCTTTTCAGGAAAGTGAA 3000
QY 3001 GGCACAAAGCACCCTCAGGCTAGCTCTATCGGACCTTCACAAGAGATGTTCCAAAGGAGATC 3060
Db 3001 GGCACAAAGCACCCTCAGGCTAGCTCTATCGGACCTTCACAAGAGATGTTCCAAAGGAGATC 3060
QY 3061 ATATGATGCTGATATGAAGTTGATGGAATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120
Db 3061 ATATGATGCTGATATGAAGTTGATGGAATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120
QY 3121 CAAACTGAATATGACAAACAGCTGGTAAATTTAATGGACTACTATGGCATATAAAGACA 3180
Db 3121 CAAACTGAATATGACAAACAGCTGGTAAATTTAATGGACTACTATGGCATATAAAGACA 3180
QY 3181 GGCTGAATATCTTGTGCTGGCATTTAAGAGCATCAAAAACCTTTTGACCGCAGAAAAGA 3240
Db 3181 GGCTGAATATCTTGTGCTGGCATTTAAGAGCATCAAAAACCTTTTGACCGCAGAAAAGA 3240
QY 3241 TGCTGAGGCCATTTAGTGTGCTGTGAGGGCTTGAGGAAGAGGCAAGAGCTTGTTCAA 3300
Db 3241 TGCTGAGGCCATTTAGTGTGCTGTGAGGGCTTGAGGAAGAGGCAAGAGCTTGTTCAA 3300
QY 3301 GAGCGCTAATGATATAGATGACATGTTACCAAGGCTTGGCTTGGTACCACGTTACATA 3360
Db 3301 GAGCGCTAATGATATAGATGACATGTTACCAAGGCTTGGCTTGGTACCACGTTACATA 3360
QY 3361 TCATCTACATATTGGGCTTGCTACAATCAGGGCTTGAAGAAGAGCTCATTTTCATTAGCTT 3420
Db 3361 TCATCTACATATTGGGCTTGCTACAATCAGGGCTTGAAGAAGAGCTCATTTTCATTAGCTT 3420
QY 3421 TCCCTGTGCTGTTTATGACCAAGCTTAATCCAGATTAAAGAGCAAAAGCAGTAAACAGGCC 3480
Db 3421 TCCCTGTGCTGTTTATGACCAAGCTTAATCCAGATTAAAGAGCAAAAGCAGTAAACAGGCC 3480
QY 3481 AGTTCTCAACTTGTCTCATCTCAGGGCTCAACTGAGTCACAGATTAGTGTGAAATGAGA 3540
Db 3481 AGTTCTCAACTTGTCTCATCTCAGGGCTCAACTGAGTCACAGATTAGTGTGAAATGAGA 3540
QY 3541 TTCAGTCGAGCGTTAAGCTGATATATATATATATATATATATATATATATATATATATAT 3600
Db 3541 TTCAGTCGAGCGTTAAGCTGATATATATATATATATATATATATATATATATATATATAT 3600
QY 3601 TGTATGCTATGTTGACTACCTTTTGTCTTTTAAACTGCAATGAAGCTGCAACATATATGC 3660
Db 3601 TGTATGCTATGTTGACTACCTTTTGTCTTTTAAACTGCAATGAAGCTGCAACATATATGC 3660
QY 3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTAATATATATATATATATATATATATAT 3720
Db 3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTAATATATATATATATATATATATATAT 3720
QY 3721 TTTCTTCTAAA 3731
Db 3721 TTTCTTCTAAA 3731

RESULT 2
AAS17837
ID AAS17837 standard; cDNA; 3731 BP.
XX
AC AAS17837;
XX
DT 08-MAY-2002 (first entry)
XX
DE Tomato RNA-directed RNA polymerase (RdRP) cDNA.
XX
```

KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
cytostatic; viricide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection;
KW immunoprecipitation; immunolocalisation; ss; gene therapy.
XX
OS Lycopersicon esculentum.

XX Key Location/Qualifiers
FH CDS 194..3538
FT /*tag= a
FT /product= "RdRP protein"
FT /EC_number= "2.7.7.48"

PN US2001023067-A1.

XX
XX PD 20-SEP-2001.

XX 08-FEB-2001; 2001US-0782874.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASSENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-595798/67.

DR P-PSDB; AAU10006.

XX New nucleic acid molecule encoding a polypeptide having the enzymatic
PT activity of RNA-directed RNA polymerase, for modulating gene expression
PT and treating cancer and virus infection in human and animals -
XX

PS Claim 1; Page 15-19; 34pp; English.

XX This sequence represents a cDNA encoding the tomato RNA-directed RNA
CC polymerase (RdRP) protein of the invention. The invention comprises the
CC nucleic acid and protein sequences of RdRP. The protein of the invention
CC can catalyze in vitro transcription of short single stranded RNAs into
CC DNA molecules, this transcription can be either primed by RNA or DNA
CC oligonucleotides or be unprimed. The protein may have cytotostatic or
CC virucide activities. The sequences of the invention may be used in gene
CC therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA
CC sequence and a template nucleic acid molecule derived from a nucleic
CC acid molecule which causes a disease are useful for treating a disease
CC caused by the undesired expression or overexpression of a nucleic acid
CC molecule in a human, rat or mouse, by administering the molecules. This
CC system can be used in the preparation of a pharmaceutical composition
CC and for inhibiting expression of any desired gene by transferring the
CC RdRP system to organisms that either lack a comparable mechanism or do
CC not sufficiently express their own RdRP. An antibody or an antagonist or
CC inhibitor to the protein are useful for inhibiting RNA directed RNA
CC synthesis and for ensuring stable heterologous, gene expression in
CC transgenic organisms. The sequence is useful for probes and/or for the
CC control of gene expression, as primers for amplification of nucleic acid
CC molecules and as tools for the detection of expression of the cDNA
CC molecules. Additionally, nucleotide and protein sequences are useful for
CC suppression of undesired gene expression in humans and animals. The RdRP
CC is useful as a therapeutic agent for the control of cancer and virus
CC infection in humans and animals and the antibody is useful for
CC immunoprecipitation or immunolocalisation of the protein. Identification
CC of polypeptides interacting with it and screening expression libraries.

XX Sequence 3731 BP; 1064 A; 569 C; 849 G; 1149 T; 0 other;

Query Match 100.0%; Score 3731; DB 23; Length 3731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATATCTTTACTTACACAGGATGACATCATCTCCCTCAAGTCTTTGTGT 60
|||||
Db 1 GAAATATCTTTACTTACTTACACAGGATGACATCATCTCCCTCAAGTCTTTGTGT 60

QY 61 GTTGTGATATAAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGTAGTTTTAT 120
Db 61 GTTGTGATATAAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGTAGTTTTAT 120
QY 121 TTTGCATAACTTCAGGGGGTATTCCAGTTGGTGTAGCATTTTGAAGTGAAGTGCACAT 180
Db 121 TTTGCATAACTTCAGGGGGTATTCCAGTTGGTGTAGCATTTTGAAGTGAAGTGCACAT 180
QY 181 GGAATTTGGTCATATGCGAAAGACAATTCAGTTTTCGGATTCCTTACTCTCTCTGC 240
Db 181 GGAATTTGGTCATATGCGAAAGACAATTCAGTTTTCGGATTCCTTACTCTCTCTGC 240
QY 241 GGAAGTGGTTAAGTCAATTTAGAGAAATATACAGATATGGAACATGTATGTCATTTGA 300
Db 241 GGAAGTGGTTAAGTCAATTTAGAGAAATATACAGATATGGAACATGTATGTCATTTGA 300
QY 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCAATTTGCCAAGTTCATTTGCCGACACAT 360
Db 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCAATTTGCCAAGTTCATTTGCCGACACAT 360
QY 361 AAGTGCTGACAAATCATCATCTTTGGCTAATAACAGCTGTATTTTGGCTCTCTCTATT 420
Db 361 AAGTGCTGACAAATCATCATCTTTGGCTAATAACAGCTGTATTTTGGCTCTCTCTATT 420
QY 421 GAAGGCTTGGGAAATGAAAACTGATATTTCCAACTCCGGGCATATGTGATCAGATGGA 480
Db 421 GAAGGCTTGGGAAATGAAAACTGATATTTCCAACTCCGGGCATATGTGATCAGATGGA 480
QY 481 TGGCATAACTTTGAAATTCGGATGTCAGATATCAGATGACAAAGTTTGCAGTGTGGGAAG 540
Db 481 TGGCATAACTTTGAAATTCGGATGTCAGATATCAGATGACAAAGTTTGCAGTGTGGGAAG 540
QY 541 TAGAGAAGTTTCAATTCATTTGGCATTTGATTGAAGAAATTTTTTTCTTTTATCTAG 600
Db 541 TAGAGAAGTTTCAATTCATTTGGCATTTGATTGAAGAAATTTTTTTCTTTTATCTAG 600
QY 601 TGGTTCAGTCACTATAAATTCAGTTTCAGTTTCATATGAAATATATGCGAGTGTGCTCCA 660
Db 601 TGGTTCAGTCACTATAAATTCAGTTTCAGTTTCATATGAAATATATGCGAGTGTGCTCCA 660
QY 661 TCGTCCATATGGTCAAAATGCTCAGTTTCCCTCATACAGTTATTTGGTGTCTCTCGGAT 720
Db 661 TCGTCCATATGGTCAAAATGCTCAGTTTCCCTCATACAGTTATTTGGTGTCTCTCGGAT 720
QY 721 CTATAAGAGACTTGAAGAACTCTGTGTATAGTCTTTAAGAAACCTCTCATATCAGATG 780
Db 721 CTATAAGAGACTTGAAGAACTCTGTGTATAGTCTTTAAGAAACCTCTCATATCAGATG 780
QY 781 GGTGAGCACACAGATTTCCCTCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA 840
Db 781 GGTGAGCACACAGATTTCCCTCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA 840
QY 841 GTTCCGTAGGGGTGTCGTCTTCCAAATTTTCAGGAAAGTTTTTCCACTATGCAGAACG 900
Db 841 GTTCCGTAGGGGTGTCGTCTTCCAAATTTTCAGGAAAGTTTTTCCACTATGCAGAACG 900
QY 901 TGAACAATATTAATTTACAGACTGTTTCACTTTTTCGTTCTCTCAAAAATTCGGTCT 960
Db 901 TGAACAATATTAATTTACAGACTGTTTCACTTTTTCGTTCTCTCAAAAATTCGGTCT 960
QY 961 GGTTCCTCAATGTCCAGCTCCCGAAGGAATTTCAATTCCTACAGATTTTGTTCAAAT 1020
Db 961 GGTTCCTCAATGTCCAGCTCCCGAAGGAATTTCAATTCCTACAGATTTTGTTCAAAT 1020
QY 1021 TAGTTCCTTTGGTACAGCATGGATGATACCTGGCCAGCATTAATGCTACTTTTCCG 1080
Db 1021 TAGTTCCTTTGGTACAGCATGGATGATACCTGGCCAGCATTAATGCTACTTTTCCG 1080
QY 1081 ATTAGTTGATCTCCGAAGGAATGTGGCATGCTTGGCATGCTTGGCATGCTTGGCATGCT 1140
Db 1081 ATTAGTTGATCTCCGAAGGAATGTGGCATGCTTGGCATGCTTGGCATGCTTGGCATGCT 1140
QY 1141 CTATATAAGAGAGTGTGTTATGATCCCGTAGGTGGCTCACTGAGCAGTATGATGGTA 1200

Db 1141 CTAATAAAGGAGTGTCTTAATGATCCGCTGAGTGGCTCACTGAGCAGTATGATGGTA 1200
Qy TCTCAAGGGTAGACAACCTCCAAATCTCCGTCCATCACITTAGATGATGGTGTGGTA 1260
Db 1201 TCTCAAGGGTAGACAACCTCCAAATCTCCGTCCATCACITTAGATGATGGTGTGGTA 1260
Qy TGTAAAGAGGTCCTAGTAACACACATCGAAGTGTATTTTGTGGTCCAGAGGTAAATGT 1320
Db 1261 TGTAAAGAGGTCCTAGTAACACACATCGAAGTGTATTTTGTGGTCCAGAGGTAAATGT 1320
Qy TTCCAATCGGGTCTCCGCAATATTTCTGAAGACATAGATAAATTTCTCGTGTTCCTTT 1380
Db 1321 TTCCAATCGGGTCTCCGCAATATTTCTGAAGACATAGATAAATTTCTCGTGTTCCTTT 1380
Qy TGTGTATGAGGAGTGGGAGAACTGTATCTACAGACTTATTAACAAAGCAAGTACTGG 1440
Db 1381 TGTGTATGAGGAGTGGGAGAACTGTATCTACAGACTTATTAACAAAGCAAGTACTGG 1440
Qy AAGTGGTGTAGGACAAACATCTATGAGAGGATCTTCAACTCTCGGGAAGGCTTTGT 1500
Db 1441 AAGTGGTGTAGGACAAACATCTATGAGAGGATCTTCAACTCTCGGGAAGGCTTTGT 1500
Qy AATTTGTCATAAAAATTTGAATTTCTGCAATTTTCACTCGAGCCAGTTGGGGATAATC 1560
Db 1501 AATTTGTCATAAAAATTTGAATTTCTGCAATTTTCACTCGAGCCAGTTGGGGATAATC 1560
Qy AGTGGGATGTTGCATCAAGACCTGGCTTACTTACAAATGATATAGAGCTTGGATGGG 1620
Db 1561 AGTGGGATGTTGCATCAAGACCTGGCTTACTTACAAATGATATAGAGCTTGGATGGG 1620
Qy TGAATTTTCCGAGATCAAGAAATGCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680
Db 1621 TGAATTTTCCGAGATCAAGAAATGCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680
Qy TTCCTCCAGAGACTTTGAGTGTCTTAGCATGAGATTCAACTTATCCCGATGATAA 1740
Db 1681 TTCCTCCAGAGACTTTGAGTGTCTTAGCATGAGATTCAACTTATCCCGATGATAA 1740
Qy GGTTCATGGAACCACTATGCTTTCTGATGGAATTGGTAAAAATATCTGTGACTTTGC 1800
Db 1741 GGTTCATGGAACCACTATGCTTTCTGATGGAATTGGTAAAAATATCTGTGACTTTGC 1800
Qy TCAATAGATGGCTCAAAATGTCGCTTCAATATACCCCATCTGCTTCCAGATTCGTTA 1860
Db 1801 TCAATAGATGGCTCAAAATGTCGCTTCAATATACCCCATCTGCTTCCAGATTCGTTA 1860
Qy TGGTGGATATAAAGGTGTGGGTGTTGATCCGGATTCATCAATGAAGTGTCTTTGAG 1920
Db 1861 TGGTGGATATAAAGGTGTGGGTGTTGATCCGGATTCATCAATGAAGTGTCTTTGAG 1920
Qy AAAGAGCATGTGGAATATGAATCAGACAACATAAAGTTAGATGTCTTGGATGGAGCA 1980
Db 1921 AAAGAGCATGTGGAATATGAATCAGACAACATAAAGTTAGATGTCTTGGATGGAGCA 1980
Qy ATATCAGCCTTGTATCTTAATCGTCAACTGATTACGCTCTTGCTACACTTGGAGTAA 2040
Db 1981 ATATCAGCCTTGTATCTTAATCGTCAACTGATTACGCTCTTGCTACACTTGGAGTAA 2040
Qy AGATGAAGTTCTCGAACAGAGCAAAAGGAAGCTGTAGATCAGCTTGTATCTTCTGCA 2100
Db 2041 AGATGAAGTTCTCGAACAGAGCAAAAGGAAGCTGTAGATCAGCTTGTATCTTCTGCA 2100
Qy TGATCTTTTGAAGGCACAGGAGGCTTTGGAATGATGTCCTCTGGAGAGACACTTAATAT 2160
Db 2101 TGATCTTTTGAAGGCACAGGAGGCTTTGGAATGATGTCCTCTGGAGAGACACTTAATAT 2160
Qy TCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTCATGCTGAGCCCTTTCTTTCAATGAT 2220
Db 2161 TCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTCATGCTGAGCCCTTTCTTTCAATGAT 2220
Qy GTTGAACACCTTCCGCGCATCCAAAGTTGCTCGAATTTGGGACTAGATCAAGAATATTTAT 2280
Db 2281 GTTGAACACCTTCCGCGCATCCAAAGTTGCTCGAATTTGGGACTAGATCAAGAATATTTAT 2280

Qy TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGATATGGTCA 2340
Db 2281 TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGATATGGTCA 2340
Qy GGTGTTTGTTCAGTTTACTGTTGCTGACATGAGAGTGTTCGTGAGATTTTACATCCATT 2400
Db 2341 GGTGTTTGTTCAGTTTACTGTTGCTGACATGAGAGTGTTCGTGAGATTTTACATCCATT 2400
Qy TATAACAGCAGATCCACCAACAGTAGTAATTTCAATCTGAAGGGAAATGTGGTGTTCGAAA 2460
Db 2401 TATAACAGCAGATCCACCAACAGTAGTAATTTCAATCTGAAGGGAAATGTGGTGTTCGAAA 2460
Qy AATTCATGCTTCGATCCTGGTGTATTTGCTGTTTAAAGGCTGTAATGTTTCGAGCGCT 2520
Db 2461 AATTCATGCTTCGATCCTGGTGTATTTGCTGTTTAAAGGCTGTAATGTTTCGAGCGCT 2520
Qy GCACCATGATGATGTTGTTGTTATTCCTCAGAAGGAAAGAACCTCATCCGAATGA 2580
Db 2521 GCACCATGATGATGTTGTTGTTATTCCTCAGAAGGAAAGAACCTCATCCGAATGA 2580
Qy ATGTTCTGGAGTGTATTTGATGGGATATCTACTTTGTTTGTGGATCAGACATGAT 2640
Db 2581 ATGTTCTGGAGTGTATTTGATGGGATATCTACTTTGTTTGTGGATCAGACATGAT 2640
Qy CCGCCAAAGCAAGTCCAGCCGATGGAATATCCTCAGACACCAGCATACAGTTGGACCA 2700
Db 2641 CCGCCAAAGCAAGTCCAGCCGATGGAATATCCTCAGACACCAGCATACAGTTGGACCA 2700
Qy TGAATGCACAAATGAGAGTGAAGTGAAGTACTTCCACCACTATATTTGTAATGACAGTTT 2760
Db 2701 TGAATGCACAAATGAGAGTGAAGTGAAGTACTTCCACCACTATATTTGTAATGACAGTTT 2760
Qy GGAATCATAGCAAAATGCCATGCTGTTATTTGAGACAGAGAACCTGATGATGCCATGAG 2820
Db 2761 GGAATCATAGCAAAATGCCATGCTGTTATTTGAGACAGAGAACCTGATGATGCCATGAG 2820
Qy TGAATGCACAAAAACTTGTCTGAGTCTTTTCAATGAGTGGACTTTCCAAAGACTGG 2880
Db 2821 TGAATGCACAAAAACTTGTCTGAGTCTTTTCAATGAGTGGACTTTCCAAAGACTGG 2880
Qy TGTTCGCCGTGAATACCATCTCAGTTGCGCCCTCAAGATACCCAGACTTTCATGGATAA 2940
Db 2881 TGTTCGCCGTGAATACCATCTCAGTTGCGCCCTCAAGATACCCAGACTTTCATGGATAA 2940
Qy GCGGCAACAGCAGCTATATCTCAGAAAGAGTATTTGGAAGCTTTTCAGGAAAGTAA 3000
Db 2941 GCGGCAACAGCAGCTATATCTCAGAAAGAGTATTTGGAAGCTTTTCAGGAAAGTAA 3000
Qy GGACAAAGCACCTCAGCTAGCTCTATCGGACCTTCACAAAGAGATGTTTCAAGGAGATC 3060
Db 3001 GGACAAAGCACCTCAGCTAGCTCTATCGGACCTTCACAAAGAGATGTTTCAAGGAGATC 3060
Qy ATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTTGACGAAGCTTTTGACTA 3120
Db 3061 ATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTTGACGAAGCTTTTGACTA 3120
Qy CAAACTGAATATGACAAACAGCTGGTAAATTTAATGGACTACTATGGCATATAAAGAGA 3180
Db 3121 CAAACTGAATATGACAAACAGCTGGTAAATTTAATGGACTACTATGGCATATAAAGAGA 3180
Qy GGCTGAATATCTTAGTGGTGGCATTTAAGAGCATCAAAAACCTTTTACCGCAGAAAAGA 3240
Db 3181 GGCTGAATATCTTAGTGGTGGCATTTAAGAGCATCAAAAACCTTTTACCGCAGAAAAGA 3240
Qy TGCTGAGGCCATATAGTGTGAGGGCTTGAAGAGGAGGCAAGAGCCCTGGTTCAA 3300
Db 3241 TGCTGAGGCCATATAGTGTGAGGGCTTGAAGAGGAGGCAAGAGCCCTGGTTCAA 3300
Qy GAGCGTAAATGATATAGATGACATGTTACCAAGGCTTTCGGCTTGGTACACGTTTACATA 3360
Db 3301 GAGCGTAAATGATATAGATGACATGTTACCAAGGCTTTCGGCTTGGTACACGTTTACATA 3360

Db 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCATTTGCCAAAGTTCAAATTTGCCGACACAT 360
QY 361 AAGTCTGACAAAATCATCACTTTGGCTAATAAACAGGCTGTATTTTGGCTCTCTCTTAATTT 420
Db 361 AAGTCTGACAAAATCATCACTTTGGCTAATAAACAGGCTGTATTTTGGCTCTCTCTTAATTT 420
QY 421 GAAGCCTTGGGAAATGAAAACGTGATTTGTCACACTGGGSCATATGTGGATCAGATGGA 480
Db 421 GAAGCCTTGGGAAATGAAAACGTGATTTGTCACACTGGGSCATATGTGGATCAGATGGA 480
QY 481 TGGCATAACTTTGAATTTGCGATGTCAGATATCAGATGACAAGTTTGCAGTTGGGGAAG 540
Db 481 TGGCATAACTTTGAATTTGCGATGTCAGATATCAGATGACAAGTTTGCAGTTGGGGAAG 540
QY 541 TACAGAAGTTTCAATCAATTTGGCATTGGATTTGAAGAAATTTTTCCTTTTATATATAG 600
Db 541 TACAGAAGTTTCAATCAATTTGGCATTGGATTTGAAGAAATTTTTCCTTTTATATATAG 600
QY 601 TGGTTCAGTCACTATAAATTCAGCTTTCATATGAAAATATATGGCAGGTTGTGCTCCA 660
Db 601 TGGTTCAGTCACTATAAATTCAGCTTTCATATGAAAATATATGGCAGGTTGTGCTCCA 660
QY 661 TGGTCCATATGGTCAAAATGCTCAGTTTCTCTCATACAGTTATTTGGTGTCTCTCGGAT 720
Db 661 TGGTCCATATGGTCAAAATGCTCAGTTTCTCTCATACAGTTATTTGGTGTCTCTCGGAT 720
QY 721 CTATAGAGACTTGAAGAACTCTGTATAGCTTCTTTAAGGAAACTCTGTATGATCAGTG 780
Db 721 CTATAGAGACTTGAAGAACTCTGTATAGCTTCTTTAAGGAAACTCTGTATGATCAGTG 780
QY 781 GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGGTATCTTCTAGCTTATTTTGGCA 840
Db 781 GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGGTATCTTCTAGCTTATTTTGGCA 840
QY 841 GTTCCGTAGGGGTGTTCTCTTCCAAATTTTCAGGAAAGTTTTCACATATGAGCAAGC 900
Db 841 GTTCCGTAGGGGTGTTCTCTTCCAAATTTTCAGGAAAGTTTTCACATATGAGCAAGC 900
QY 901 TGAAGCAATATTAATTTACAGACTGTTTTCACCTTTTTCGTCCTCAAAATCGGCTCT 960
Db 901 TGAAGCAATATTAATTTACAGACTGTTTTCACCTTTTTCGTCCTCAAAATCGGCTCT 960
QY 961 GGTCCCAATGTCAGCTCCGGAAGAAATTTCAATTCCTTACAGATTTTGTTCAAAT 1020
Db 961 GGTCCCAATGTCAGCTCCGGAAGAAATTTCAATTCCTTACAGATTTTGTTCAAAT 1020
QY 1021 TAGTTCTTTGGTACAGCATGATGCATACCTGGCCAGCATTAATGTCTACTTTTCCG 1080
Db 1021 TAGTTCTTTGGTACAGCATGATGCATACCTGGCCAGCATTAATGTCTACTTTTCCG 1080
QY 1081 ATTAGTTGATCCTCGAAGGAGAAATGTGGCATGCAATGAGCATGCCCTTAGAGAACTGTA 1140
Db 1081 ATTAGTTGATCCTCGAAGGAGAAATGTGGCATGCAATGAGCATGCCCTTAGAGAACTGTA 1140
QY 1141 CTATATAAGGAGTCTGTTATGATCCGCTGAGTGGCTCACTGAGCAGTATGATGGTA 1200
Db 1141 CTATATAAGGAGTCTGTTATGATCCGCTGAGTGGCTCACTGAGCAGTATGATGGTA 1200
QY 1201 TCTCAAGGATAGACAACTCCAAATCTCCGTCATCACTTAGATGATGGGTTGGTGA 1260
Db 1201 TCTCAAGGATAGACAACTCCAAATCTCCGTCATCACTTAGATGATGGGTTGGTGA 1260
QY 1261 TGTAGAAGGTTCTTAGTAAACCAATGCAAGTTTATTTTGGTGGTCCAGAGGTTAATGT 1320
Db 1261 TGTAGAAGGTTCTTAGTAAACCAATGCAAGTTTATTTTGGTGGTCCAGAGGTTAATGT 1320
QY 1321 TTCCAAATCGGTTCTCCCAATTTCTGGAAGACATAGATTAATTTCTCTGTTTCTCTT 1380
Db 1321 TTCCAAATCGGTTCTCCCAATTTCTGGAAGACATAGATTAATTTCTCTGTTTCTCTT 1380
QY 1381 TGTGATGAGGAGTGGGAGAACTGTATCTACAGCTTATTAACCAAGCAAGTACTGG 1440
Db 1381 TGTGATGAGGAGTGGGAGAACTGTATCTACAGCTTATTAACCAAGCAAGTACTGG 1440

QY 1441 AAGTGTGTGAGGACAAAACATCTATGAGAGATCTTATCAACTCTCGGAAAGCCTTTGT 1500
Db 1441 AAGTGTGTGAGGACAAAACATCTATGAGAGATCTTATCAACTCTCGGAAAGCCTTTGT 1500
QY 1501 AATTTGGTGAATAAAATTTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGGATATTC 1560
Db 1501 AATTTGGTGAATAAAATTTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGGATATTC 1560
QY 1561 AGTGTGATGTTTGGCATCAAGACCTGGCCCTTACTGCAATATGATATAGAGCTTGGATGG 1620
Db 1561 AGTGTGATGTTTGGCATCAAGACCTGGCCCTTACTGCAATATGATATAGAGCTTGGATGG 1620
QY 1621 TGATTTTCCAGATCAAGAAATGTCGCAAAATATGTCGAGACTTGGTCAATCTTTTGG 1680
Db 1621 TGATTTTCCAGATCAAGAAATGTCGCAAAATATGTCGAGACTTGGTCAATCTTTTGG 1680
QY 1681 TTCTCTCAGAGAGACTTTTGAAGTTTCTTAGGCATGAGATTGAAGTTATTTCCGATGATA 1740
Db 1681 TTCTCTCAGAGAGACTTTTGAAGTTTCTTAGGCATGAGATTGAAGTTATTTCCGATGATA 1740
QY 1741 GGTTCATGGAACAGCTATGCTTTTCTGATGGAATTTGTAATATCTGGTGCATTTGC 1800
Db 1741 GGTTCATGGAACAGCTATGCTTTTCTGATGGAATTTGTAATATCTGGTGCATTTGC 1800
QY 1801 TCATAGAGTTGCTCAAAATGTCGCTTCAATATACCCATCTGCTTCCAGATTCGTTA 1860
Db 1801 TCATAGAGTTGCTCAAAATGTCGCTTCAATATACCCATCTGCTTCCAGATTCGTTA 1860
QY 1861 TGTGTGATATAAGGTTTGTGCTGCTTCAATCGGATTCATCAATGAAGTTGCTTTGAG 1920
Db 1861 TGTGTGATATAAGGTTTGTGCTGCTTCAATCGGATTCATCAATGAAGTTGCTTTGAG 1920
QY 1921 AAGAGCATCTCGAATATGAATCAGACACATAAAGTTAGATGCTCTTGGATGGAGCAA 1980
Db 1921 AAGAGCATCTCGAATATGAATCAGACACATAAAGTTAGATGCTCTTGGATGGAGCAA 1980
QY 1981 ATATCAGCTTGTATCTTAACTGCTCAACTGATTACTCTTCTGCTACACTTGGAGTGA 2040
Db 1981 ATATCAGCTTGTATCTTAACTGCTCAACTGATTACTCTTCTGCTACACTTGGAGTGA 2040
QY 2041 AGATGAAGTTCTCGACAGAGCAAGAAAGAGCTGTAGATCAGCTTGTATCTTGA 2100
Db 2041 AGATGAAGTTCTCGACAGAGCAAGAAAGAGCTGTAGATCAGCTTGTATCTTGA 2100
QY 2101 TGATTTCTTCAAGGACACAGGAGCTTTGGAATTTGATGCTCTCTGAGAGAACTAATAT 2160
Db 2101 TGATTTCTTCAAGGACACAGGAGCTTTGGAATTTGATGCTCTCTGAGAGAACTAATAT 2160
QY 2161 TCTCAAGGCAATGCTAACTGCTGATTAAGCTGATGCTGAGCCCTTCTTCAATGAT 2220
Db 2161 TCTCAAGGCAATGCTAACTGCTGATTAAGCTGATGCTGAGCCCTTCTTCAATGAT 2220
QY 2221 GTTGCAAACTTCCGCGCATCCAGTTGCTGATTTGCGGACTAGATCAAGAAATATTTAT 2280
Db 2221 GTTGCAAACTTCCGCGCATCCAGTTGCTGATTTGCGGACTAGATCAAGAAATATTTAT 2280
QY 2281 TCCAAATGGAAGAACATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340
Db 2281 TCCAAATGGAAGAACATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340
QY 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGTTTCTGAGAGTTTACATCCATT 2400
Db 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGTTTCTGAGAGTTTACATCCATT 2400
QY 2401 TAATAACAGCAGATCCCAACAGTAAATTTCAATTTGAAGGAAATGTGGTGTGTCAAA 2460
Db 2401 TAATAACAGCAGATCCCAACAGTAAATTTCAATTTGAAGGAAATGTGGTGTGTCAAA 2460
QY 2461 AATCCAGTCTGCAATCTGCTGATATTCGTTGTTTAAAGGCTGAATGCTCGAGGCT 2520
Db 2461 AATCCAGTCTGCAATCTGCTGATATTCGTTGTTTAAAGGCTGAATGCTCGAGGCT 2520

318	AAAGGTGGGTTC	CCAAAGACGATTT	GC	AAATTAATTC	AAATTC	CAACACCGCAAAATCT	CGCTACA	377
371	AAAATCATCACT	TTGGCTAATAAC	-----	AGCGTGTAT	TTTTGGCTCTCTCTATT	TTG	421	
378	TCATGATGTC	AGAGCTAAC	CAATTTTG	AACAATTTTG	AACAATTTTG	CGGTATG	437	
422	AAAGCTTGGAA	TGAATAAACT	GATATTG	TCCAACTGC	--GGGCATAT	TGGATCAGATG	478	
438	AAAGCTCGGGA	ATGAAAGAGAT	ATGTC	CAAGCCCAAGGGT	TTTTTTC	GCATAGTTTG	497	
479	GATGGCATAC	ATTGTAATTT	TCGGATGTC	CAGATATC	AGATGACAAGTT	TCAGTGT	535	
498	GATGATGAA	CTGCTTTT	GSCGT	CTCAGATCT	CAAGGGAAGNAT	CTCTGTTTATG	557	
536	GGAAATACAGA	AGTTTCAATTC	AAATTTGG	CAATGATG	GAAGAAATTT	TTTTTTCTTTTA	595	
558	AAAAAGCAGG	ATGTTATTTAA	ATTTTGG	AGTGGGAAT	TGAAGAAAGAT	GCATTTCTTATTT	617	
596	TCATAGTGG	TTACGCTACAT	AAACTTTC	AGCTTTTC	ATATGAAATAT	ATATGCGAGTTGTG	655	
618	TCCCAACAACA	ATGTGCAAT	TACAACCT	TGACCTTT	TCATATG	AGACATTTTGAAGATTTGAG	677	
656	CTCCATGCT	CCATATAGT	CAAAATG	CTCAGTTT	CTCTCATAC	AGTTATTTTGGTCTCCT	715	
678	CTGCATCGG	CCACGGAAT	GAGACT	ACACGTTAT	CTCTGATT	TCAGTTACTTTGGTCTCC	737	
716	CGNATCTAT	TAAGAC	-----	-----	-----	TTGAAACTCTCTGTATAGC	751	
738	CGGGTTTTT	GGAAACGAT	GTACCT	ACATCA	CAAAATATCT	TTTGATGAPCCTTTGTTCAAC	797	
752	TTCTTTAAG	AAAAACT	CCCTGAT	GATCAGT	GGGTGAGGACAA	CAGATTTCCCTCC	808	
798	TTCTTCAAGA	ATCCCTGAT	GAGCAAT	GGATCC	GAGCAATGAT	TTTCACTCCAGAAAGT	857	
809	TGGATAGG	GTATCTTCT	TAGCTTAT	TTTGG	CAGTTTCCGT	TAGGGTGTCTGCTTTCCAAAT	868	
858	CGTATTGGG	CAGTCCCT	CCGCATAT	GTCTG	AGCTTCCTTAAT	TGCGCCGACAACTTTCCAAAT	917	
869	TTGAGGAA	AGTTTTT	TCCACAT	ATG	CAGAACCTG	AAAAACAATATTTAC	TTTACAGACTGGT	928
918	TTCAGGAA	AAACTTT	CGTTAT	TATGAG	AAACTGAGAGCAAT	ACACITTTACACACAGGA	977	
929	TTACCTTTT	TCCTCTCT	CAAAAAT	CGGCTCT	GGTTTCC	CAATGCCAGCCTCCGAGGA	988	
978	GTTCCTCTT	-----	TCTCAAAAT	TGGGTCT	TGCCCCAT	TGTTGCTCTCTCTAGT	1031	
989	ATTTCAAAT	TCCTAC	AAAGATTT	TGTTTCA	AAATTAGTCT	TTTGGTACACATGGATGCAT	1048	
1032	GTAAAAAT	ATCATG	ACATCT	TGTTTAA	AGTCAATTCAT	TGGTTACACATGCAATGCTT	1091	
1049	CCTGGCC	ACGATTA	ATCTACTTTT	TCCGAT	TAGTTGAT	CTCTGAAGGAGAAATGTG	1108	
1092	GCAGGACCT	GCACCT	TGATGGT	GACTCT	ATATCGCT	TGGTTGATCCACGTFAGAAATCGCCCGT	1151	
1109	GCATGCAT	TGAGCAT	GCCTTAG	NAGAAC	TGTACTAT	ATAAGGAGTGCCTGTATGATCCC	1168	
1152	GAATTTATT	GAATAT	GCCTTTAG	AAAAAGATTT	TACTATTC	AAAGGAATTTTGGTTATGAAACC	1211	
1169	GTGAGTGG	CGTCACT	GAGCAGT	ATGATGG	TATCTCA	AGGTTAGACAACCTCCAAAATCT	1228	
1212	ACAAAGTGC	TGACTGAT	CAGTAC	AAAAAT	ATACCT	TGAGTCAAAAAATCATCCTCGGTCA	1271	
1229	CCGTCCAT	CACACTTT	AGATGAT	GGTTGGT	GTATGT	ATAGNAGGGTCTCTAGTAACACCATGC	1288	
1272	CCTGCAAT	ATCTCT	TGGATAC	AGGGTTGG	TATACGTT	TCGACGGTTTCAGATCACGCCCTGC	1331	
1289	AAAGTTTAT	TTTTTGT	GTGTC	CAGAGGTT	TAATGTTT	CCAAATCGGGTTCTCCGCAAAATTTCT	1348	
1332	AAAGTATAC	TTTTTGT	GGTCC	AGAGATG	ATGATG	CTCAATCTGTGTTCTCCGTCATTTCCGT	1391	
1349	GAAGACAT	AGATAACT	TTTCT	CGTGT	TTCTTTTGT	TGATGAGAGTGGGAGAACTGAT	1408	
1392	GAACATATT	GATAACT	TTTCT	ACGTGTTT	CATTGTTT	GTATGAGAAATTTGGATAAACTGTTT	1451	

```
QY 2477 CCTGATGATATTCGTGTTTAAAGGCTGTAATGTTTCAGCGCTGCACCATGTTAGT 2536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2511 CCAGTGATGCGGTGTTTACAGCTGTGATGTCACGATTTGTACACATGGTGGAC 2570

QY 2537 TGTGTTGATTCCTTCAGAAAGAAAGACCTCATCCGAATGAATTTCTGGGAGTAT 2596
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2571 TGTGTTTTCCTTCAAAAAGGACCAAGACCTCATCAATAGTGTTCGGGAAGTAT 2630
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2597 TTGATGGGATATCTACTTGTGTTGGATCAAGACATGATCCCGCAAGGCAAGTC 2656
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2631 CTGGATGGATATCTACTTGTGTTGGACCATGAATGATTCCTTCGCCCAATT 2690
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2657 CAGCCGATGGAATATCCTCCAGCACCCAGCATACAGTTGGACCATGATGCACAAATTGAG 2716
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2691 GATCCAATGGATATCTACTGCTCCGCAACTGTGGAATGGATCATGATGATGATCGAG 2750
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2717 GAAGTTGAAGAGTACTTCCACCACTATATGTTGAATGACAGTTTGGGAATCAGCAAAAT 2776
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2751 GAGTTGAGGAGTATTTGCCAATACATAGTCAATGACAGTCTGGGAATTAATGGCCAAAT 2810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2777 GCCCATGTCGTATTTGCAGACAGAGAACCTGATATGCCCATGAGTATCCATGCACAAAAA 2836
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2811 GCACACACTGCTTTCGAGATAAAGAACATTTGAAGCAATGCTGATCAATGTGTTAAG 2870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2837 CTTGCTGAGCTCTTTCAAITGCACTGAGTCTTCCAAAGACTGTTCCGCTGGAATA 2896
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2871 CTTGCAAGTTGTTTCAACAGCAGTTGACTTTCCTAAAAGTGTGTTCCAGCTGTTATA 2930
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2897 CCATCTCAGTTGCGCCCTTAAGATAATCCAGACTTCATGATAGCGGACAGACCCAGC 2956
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2931 CCTCCTGAACCTTATGTCACAAAGATATCCTGACTTCATGGAGAGCCTGACAAACCCACA 2990
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2957 TATATCTCAGAAAGAGTATTTGGAAGCTTTTCAGGAAGTGAAGGACAAAGACCTCAG 3016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2991 TACAATTCGCAATACGTTGATAGGAAGCTTTTAGGAAGTGAAGAAATATCAACAAGT 3050
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3017 GCTAGCTATATCGGACCTTCACAAAGAGATGTTTCAAGAGGAGATCATATGATGCTGATG 3076
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3051 GCCGCTCAATATACATCTTCACAAAGTGTGTCGAGAGACTTTAGACCATGAATG 3110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3077 GAAGTTGATGATTTGAAGTATCATTTGACAGAGCTTTTGACTACAAAACCTGAATATGAC 3136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3111 GAAGTTGATGCTTCATGGATTATGTTGATGATGCTTTCTATCACAACCAATATATGAC 3170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3137 AACAGCTGGGTAAATTAATGGACTACTATGGCATTAACAGAGGCTGAAATACTTAGT 3196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3171 TACAAGTTGGGAATCTGATGGACTACTATGGGATCAAAACTGAAGCTGAATCCTCGT 3230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3197 GGTGGCATATGAAGGCATCAAAAACCTTTTGACCGCAGAAAAGATGCTGAGGCCATAGT 3256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3231 GGGATATTTATGAAATGTCAAAATCTTCAACAAAAGGAGGGATGCAGAAGCAATCAAT 3290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3257 GTTGTGTGAGGCTTCAGGAAGGAGGCAAGAGCCTGGTTC-----AAGAGCGTAAT 3310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3291 ATGGCTGTGAGTCCCTTAGGAAGAGGCGCAGGGCTTGGTTCAATGAAAACAGCAGTGGT 3350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3311 GATATAGA-----TGACATGTTACAAAAGCTTCGGCTTGGTACCAGCTTAC 3358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3351 GATGTAGATTACGGGAGTAGTGATGTATGCAAAAGCTTCTGCTTGGTACCATGTTACT 3410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3359 TATCATCTACATATGGGTGTGTACAAATCAGGGTGTGAAGAGCTCATTCATTAGC 3418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3411 TATCATCAAGTTACTGGGTGTGTATTAATGAAGGCAATGAATGGGATCAATTATCTAAGT 3470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3419 TTTCCCTGTTGTTTATGACACACTAATCCAGATTAAGAGGACAAA 3466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3471 TTCATGTTGTTTACCCCTCTTCTTGTCCAATCAAGAAAGAGAAA 3518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 5

AAC63739

ID AAC63739 standard; cDNA; 3737 BP.

XX

```
AC AAC63739;
XX 07-FEB-2001 (first entry)
DT
XX
DE Maize RNA-directed RNA polymerase EST sequence, SEQ ID NO: 5.
XX
KW Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;
KW gene mapping; gene marker; plant virus resistance; plant breeding; ss.
XX
OS Zea mays.
XX
PN WO200060097-A1.
XX
PD 12-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09105.
XX
PR 07-APR-1999; 99US-0128094.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX
WPI: 2000-679376/66.
DR P-PSDB; AAB28330.
XX
PT New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance.
XX
PS Claim 2; Page 35-36; 62pp; English.
XX
CC The present sequence is one of a number of cDNA molecules which encode
CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC probes for genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes.
XX
SQ Sequence 3737 BP; 1063 A; 760 C; 863 G; 1051 T; 0 other;
```

Query Match 26.5%; Score 987.8; DB 21; Length 3737;
Best Local Similarity 59.8%; Pred. No. 7e-261;
Matches 1997; Conservative 0; Mismatches 1207; Indels 135; Gaps 15;

```
QY 195 TGGGAAGCAATCACTAGGTTTCGGATTCCCTTATCTCTCTCTCGGGAAGTGGTTAACT 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 TTGGGAGAACTATTTCAGTCCCAAGGTTTGTCTCTACTGACAGTGCAGAAATCTGCAAT 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 CATCTCTAGAGAAATATACAGGATATGGAACCTGATGTGCTATTGGAGTTAAACAGTCCA 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 TGTTTTGGAGCAATTCGTTGCTGTGGAACCATCTGCTCTCAAGCTCAGGCATCCAA 203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 315 AA-----GGAGATCTAGAGCAATTCGCAAGTTCAATTTGCCGACACATAAGTG 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 GGAACATCTCTGCCAACTCAAGGCAATTTGCTATAGTTTCAGTTCACAGTCCAGGAAAGTG 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 366 CTCACAAAATCATCACTTTGGCTAATA---ACAGGCTGTATTTTGGCTCTCTCTATTGCA 422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 CTCATTTGTAGAGATGCGGTCAAGACAGGTTCTCAGATTGGACGGTTTATCTGA 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 423 AGGCTTGGGAATGAAAACATGATATTGTCC---AACTGGGCAATATGTGGATCAGATGG 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 GAACACAGACTCGACACCGGACATTTGTCGAAGCAAGGATTCCAATGTTTCTCTAG 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 480 ATGGCATAACTTTGAATTTCCGATGTCAG---ATATCAGATGACAAGTTTGCAGTGTGG 536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 AGGACATTTGCTGCAATTTGGGATGCTTGGTTAAGAAATATCCTATCTGCTCTTTTA 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 537 GAAGTACAGAAGTTTCAATTCAATTTGGCATTGGATTGAAGAAATTTTTTCTTTTAT 596
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 444 GAGCAAGTAATGTTTCGGTTTCAATTTTGGATTGATATGAAAAAGATCTACTTCTAOCCTCT 503
QY 597 CTAGTGGTTCAGCTAGTATPAAACTTCAGCTTCATATGAAATATATGCGAGTTGCG 656
Db 504 CCTCAATTTTACTAAATTTAACTTGAACCTTCTTACGAAAGTATATGGAGATGCGAGC 563
QY 657 TCCATCGTCCATATGGTCAAAATGCT-----CASTTTCTCCCTCATACAGTTATTTGGTG 710
Db 564 TTCACGGTCCACCTGCTTATAGGTCACGGACACAGTTCCTTTTGATTCAAGTTCAGGCAG 623
QY 711 CTCCTCGGATCTATAAGAGAC-----TTGAAACTCCTGTT 746
Db 624 CTCCTAAATTTATAAACTGTCCTCCAGGCGCTCCAGGTCCTATGTTTGGAGATCCTTTCT 683
QY 747 ATAGCTTCTTTAAGGAAACTCCTGATGATCAGTGGGTGAGGACACAGATTTCCCTCCAT 806
Db 684 TCACTGGTTTAGGATGACACAGATGAACAATGGACAGGACAATTTGATTTACTCCAT 743
QY 807 ---CTTGATAGGGTATCTCTAGCTTATGTTTGGAGTCCGTAGGGGTGCTGCTTC 863
Db 744 CAGCTAGCATCGGCAATCATCTATTTATGCTGGAGGTGCCACAACAGTGTGAGCTTC 803
QY 864 CAATTTCCAGGAAAGTTTTTCCACATGACAGAACGTGAAACAATATTACTTTACAGA 923
Db 804 CAAGNAITGGCGACTACTTTGTTTACATRAAGAGCAGAACTTTGACTTTGAATGTCGA 863
QY 924 CTGPTTCACCTTTTCTGCTCTCAAAAATCGGCTCTGGTTCCTCAATATGTCAGCTCCGG 983
Db 864 ATGGGTATTATCTCTGTGTGTAGCAA-----CCTGTACCAATTTGAAATCTCCTG 917
QY 984 AAGGAATTTCAATCCCTACAGATTTTGTCAAAATTAGTCTTTTGGTACAGCATGGAT 1043
Db 918 ATTACATAGAGTCCCTATGAGATACTCTTCAAAATCAACCAATTTGGTTCAGATGGGA 977
QY 1044 GCATACCTGGGCGCAGCATTAATATGCTACTTTTCCGATTTAGTTGATCTCGAAGGAGAA 1103
Db 978 CACTCAGTGGGCGCAACAGTGTATCATAGTTTCTTCCGTCATGTTAGCCCAAAATTTGAAC 1037
QY 1104 ATGTGCGATGCTAGCATGCTTTAGAGAACTGTACTATATAAGGAGTGCCTGTATG 1163
Db 1038 CTATTGATCATATAAAGCAGCACTTTTAAGATGACATATTTGAAAGCAGCTGCTGA 1097
QY 1164 ATCCCGTGAAGTGGCTACTGAGCAGTATGATGGGTATCTCAAGGGTATAGA---ACCTC 1220
Db 1098 ACCCAACAGATGTTTACTGTGCAATACTCCAGAAATACGGAATACGCCATGCATCAC 1157
QY 1221 CAAATCTCCGTCACATCACTTTAGATGATGGTGTGCTATGTAAGAGGGTCTTAGTAA 1280
Db 1158 AAAAGTTATCTAATATATCTCTGATGATGGCTTGGTCTATGTCCACAGGGTGCAGTTA 1217
QY 1281 CACCATGCAAGTATTTTGTGTCAGAGGTTAAATGTTTCCAACTCGGGTCTCCGCA 1340
Db 1218 CCCCCTGCTAAAGTATTTTATGGACCTGAGATAAATGTCTCCAATCGCGTTGCGGC 1277
QY 1341 ATTATCTGAAGACATAGATAACTTTCTTCGTGTTCTTTTGTGATGAGGAGTGGGAGA 1400
Db 1278 ATTCTCTCGACACATAGATAACTTCTTCGGATTTCAATTTGTTGATGAAGACTGTGAGA 1337
QY 1401 AACTGATTTCTACAGACTTATACCAAAAGCAAGTACTGGAAGTGGTGTGAGGACAAACA 1460
Db 1338 AGCTCGGTTCAAGTATTTGTCACCTGATCTACTCTCGGAATGATCGAAGGAACTG 1397
QY 1461 ---TCTATGAGAGGATCTTTATCACTCTCGGAAAGGCTTTGTAATGGTGTATAAAAAT 1517
Db 1398 CTCTGTATAATAGAGTTTGTGAGTCCCTTCAATATGGCATCAATATTGGTGACAAGCACT 1457
QY 1518 TTGAATTTCTGCAATTTTCAATCGCCAGTTCGGGATAAATCAGTGTGATGTTTGGAT 1577
Db 1458 TTGAGTTTCTTGCTTTTCTCAAGTCAGCTTCGAGATAACTCTGATGGATGTTTGGTT 1517
QY 1578 CAAAGCCTGGCCTTACTGCAATATGATATAAGAGCTTGGATGGGTGATTTTTCGAGATCA 1637
Db 1518 CTCGGGAGGATGATCTGCGAGGACATAAGGAAGTGGATGGGACTTTCGAGATATCA 1577

QY 1638 AGAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTTGGTTCCTCCAGAGACAT 1697
Db 1578 GAAATGTCGCAAAATATGCTGCAATACTTTGGGCAATCTTTCACTTCTCCTCAAGAACTT 1637
QY 1698 TGAGTGTCTTAGGATGAGATGAAGTATTTCCCGATGAAAGGTTCAATGGAACAGCT 1757
Db 1638 TAAAGTACACAATCTGAGGTGGAAGAAATCTCTGATATTA---CAATGCGCAAAAGT 1694
QY 1758 ATGCTCTTTCTGATGAAATGTTGAAATATCTGGTGACITTTGCTCATAGAGTTCCTCAA 1817
Db 1695 ACATATCTCTGATGAGTGGAAAGATCTCAGCTAATTTGCGAGTGGAGTGGCTATGA 1754
QY 1818 AATGTCGCTTCAA---TATACCCCTCTGCTTTCAGATTCGTTATGTTGATATAAG 1874
Db 1755 AGTSCAAATGAAACGCTTTGCTCTTCTCTCCTCAGATGATGCGGTTTACAAAG 1814
QY 1875 GTCTGTGGGTGTGATCGGATTCATCAATGAAGTGTCTTTTGAGAAAGAGATGTCGA 1934
Db 1815 GTGTTGCTGTAGATACAGATCAATCATAAAGCTTTCTTTGAGAAAAGCATGTCMA 1874
QY 1935 AATATGAATCAGACACATAAAGTTAGATCTCTTGGATGGAGCAAAATATCAGCCTGTT 1994
Db 1875 AGTTCAGTCAGAAAATATCAGTCTTGAIGTCTCTTGCATACAGCAAGTACCAACCATGCT 1934
QY 1995 ATCTTAATCTCACTGATTAAGCTCTTGTCTACACTTGGAGTGAAGATGAAGTTCCTG 2054
Db 1935 TCTGAAATCGGAGTGTGATTTACTTCTCTCAGACTTGGGTTAGGATATGTTCTTTG 1994
QY 2055 AACAGAAGCAAAAGGAGCTGTAGATCAGCTGTATGCTATCTTGGATGATCTTTTGAAG 2114
Db 1995 AGCTAAAGCAAGAGGAGCCTTAAGGCAATTTGAACAGATTTGAACAGTGAACACACAGCTG 2054
QY 2115 CACAGGAGGCTTGGAACTGCTCTCGAGAGAACACATAATATTCTCAAGGCAATGC 2174
Db 2055 CTGCTGAAGCAGTGAATTTATGCCCATGGAGAGGTAACTAGTATTAAGAAATGT 2114
QY 2175 TAAACTGTGTTATAAGCCTGTAGCTGAGCCCTTTCTTCAATGATGTTTCAACCTTC 2234
Db 2115 TGTCTATGCTGCTACAGCCTGATCATGAGCCATATCTTTCCATGCTGCTCAAACTTTA 2174
QY 2235 GCGCATCAAGTTGCTGATTTGGGACTAGATCAAGAATATTATTTCCAAATGGAGAA 2294
Db 2175 GAGCTCCAGCTCTAGATTTGAAACAAGTCAAGGATATTCATCACACAGGGGAG 2234
QY 2295 CAATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGTCAGGTTTGTTCAGT 2354
Db 2235 CAATGATGGTTGCTGGATGAAACCTGCACACTTAAGTACGGCCAGGTTATCGTCCAG 2294
QY 2355 TTACTGTGCTGGACATGGAGGTTTCTGACGATTTACATCCATTTAATAACAGCAGAT 2414
Db 2295 CTCTTTACAGTGCAGATG----- 2312
QY 2415 CCACCAACAGTAAATTTCAATCTGAAGGAAATGTTGTTGTCAAAAAATCCATGCTTC 2474
Db 2313 ---ACCATCGCAAGTCTGTTAACTGGAAGAGTGTGTCGCCAAAAAATCTTTGTCTCC 2369
QY 2475 ATCTGTTGATPATTCGTGTTTTAAAGCTGTAAATGTCGAGCCTGCACACATGTFAG 2534
Db 2370 ACCCTGTGATCATCGGTTCTCCAGGCTGTGATGTTCTGCTCTGCACCTTGTGTTG 2429
QY 2535 ATTGTGTTGTTATCCCTCAGAAAGGAAAAAGACCTCATCCGAATGAATGTTCTGGGAGT 2594
Db 2430 ACTGTGTTGTTCTTCCACAGCAGGACCAAGCCGACCCCTAATGATGTTTCAGGGAGT 2489
QY 2595 ATTTGGATGGGATATCTACTTTGTTGCTGGGATCAAGACATGATCCCGCCAGGCAAG 2654
Db 2490 ATCTGATGGGACATATATTTTCTTGGGATCCACATCTTATTCCAAAGTCTTTGG 2549
QY 2655 TCCAGCCCATGGAATCTCTCCAGCCACAGCATACAGTGTGACCATGATGTCACATTTG 2714
Db 2550 TGGATCTTATGACTATCTACTCCAGCTTCAGCAGAAACATTTAGACCATGATGTCATTTG 2609

```
QY 2715 AGGAAGTTGAGAGTACTTCCACCAACTATATTGTGATGACAGTGTGGGAATCATAGCAA 2774
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2610 AGGAGATACAGGAGTACTTCCAAACATACATAGTAAATGAGAGTCTTGGGATATCGCCA 2669
QY 2775 ATGCCCATGTCTGATTTTGCACACAGAGAAGCTGATATGGCCATGAGTATCCCATGCAAAA 2834
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2670 ATGGCATGTGTTTACAGATCAGCAACGATGATGAAAGCTGAGAGTCCACCGTGGTTC 2729
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2835 AACTTGTGAGCTCTTTTCAATTCAGTGCAGTCTTCCAAAGACTGGTGTCCCGCTGAAA 2894
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2730 AACTGGCCAAAGCTCTTCTATAGCTGTGATTTCCCAAAGACTGGAGTCCCGCTCTGA 2789
QY 2895 TACCATCTCAGTTGCGGCCCTAAAGAAATACCCAGACTTTCATGGATAAGCCGGAACAGCCA 2954
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2790 TTCCACATGACCTACATGTCAGGAGTATCTGACTTCATGGAGAACTCGACAAAGTCA 2849
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2955 GCTATATCTCAGAAGAGTATTTGGAAGCTTTTCAGGAAGTGAAGCAAAAGCACCTC 3014
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2850 CCTATGAATCAAGAGGTGTATCGGGAAGCTCTATAGGGAATTAAGAACACACACACC 2909
QY 3015 AGGCTAGCTCTATCGGACCTTCCAAAGAGATGTTGCAAGGAGATCATATGCTGATA 3074
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2910 A-----CATAAACACTTCACGAGGAGTGGCAAGTGGCAAGCGCTCTATGACACCGATT 2960
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3075 TGAAGTTGATGGATTTGAAGATACATTTGACGAAGCTTTTGACTACAAACTGAAATATG 3134
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2961 TGATTTGATGGCTATGAAGATTACATTTACTGAGGCTATAGATTTCAAGGAAGAGTAGC 3020
QY 3135 ACAACACGCTGGTAATTTAATGACTACTATGTCATGCAAAACAGAGGCTGAAATACTTA 3194
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3021 ATTTGAGCTGGTAACTTATGACCACTATGSCATATAAAGTGAAGCTGAGATATAA 3080
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3195 GTGGTGGCAATATGAAGCATCAAAAACCTTTTGACCGCAAAAAGATGCTTGAGGCCATTA 3254
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3081 GTGGATGATTTCTAAGATGCAAAAGATTTCCAAAGAGTAGTGTGCTGATGCAATTA 3140
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3255 GTGTTGCTGTAGGCCCTTGAGGAGGAGGCAAGAGCTGTTCAAGAGCGCTA----- 3308
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3141 GAATGCGGTGAGATCTTTGAGGAAGAAGCTAGGTGCTGTTCAATGAGATGAGCAGACAG 3200
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3309 -----ATGATATAGATGACATGTTTACCAAAGGGCTTCGGCTTGATACCATCTTACATATC 3362
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3201 GAGAGGATGGCAAGATGCCATGGAGGCCAAGGCTCTGCTTGATACCATCTTACTTATC 3260
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3363 ATCCATACATATTTGGGGTGTACATATCAGGGGTGAAAGAGCTCATTTCAATGAGCTTTC 3422
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3261 ATCAGCAGTACTTGGGCGAGCTACATATGAGGGTATGATCGGCCGATCTTTATGCTTCC 3320
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3423 CTGTGTGTGTTTATGACAGCTAATCCAGATTAAGAGG 3461
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3321 CATGTGGCTATGCAAGCTTGTGGCCATCAGCAGG 3359
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
AAC63740
ID AAC63740 standard; cDNA; 3901 BP.
XX
AC AAC63740;
XX
XX 07-FEB-2001 (first entry)
XX
DE Maize RNA-directed RNA polymerase FIS sequence, SEQ ID NO: 7.
XX
KW Maize; plant; RNA-directed RNA polymerase; FIS; gene mapping;
KW gene marker; plant virus resistance; plant breeding; ss.
XX
OS Zea mays.
XX
XX WO2000060097-A1.
XX
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09105.
XX
PF
```

```
XX 07-APR-1999; 99US-0128094.
PR
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
PI WPI: 2000-679376/66.
XX P-PSDB; AMB28531.
DR
DR New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX PT for controlling gene expression and providing mechanisms to engineer
XX PT plant virus resistance
XX
XX Claim 2; Page 40-41; 62pp; English.
XX
XX The present sequence is one of a number of cDNA molecules which encode
CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC probes for genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes.
XX
XX Sequence 3901 BP; 928 A; 975 C; 1058 G; 935 T; 5 other;
SQ
Query Match 9.2%; Score 345; DB 21; Length 3901;
Best Local Similarity 52.4%; Pred. No. 5.4e-84;
Matches 1049; Conservative 0; Mismatches 900; Indels 54; Gaps 11;
QY 1223 AATCTCCGTCCTACTTATAGATGATGGTGTGTTATGTAAGAGGTCTCTAGTAACA 1282
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1290 AAACATCTTCGCAACAGAGATGGGGATGCAATTCCTGAGTGAGGAGCTGTATCAAG 1349
QY 1283 CCATGCAAAAGTTTATTTTGTGTCAGAGAGTTAATTTTCCAAATCGGGTTCTCCCAAT 1342
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1350 CCCACCAAGCTTATTGTCATGCCACCAGAGTGGAGCGCTCTAATCGCTCATCCGCGAT 1409
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1343 TATTTCTGAACATACATACATTTCTTCGTTTCTTTTCTTTTGTGATGAGGAGTGGAGA-- 1400
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1410 TATAGTGAAGTCTCAGACCGGTTTCTGAGGGTTACTTTTATGGATGGGAATGAGATG 1469
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1401 ---AAGTGTATTTCTAGAGACTTATTACCAAAAGCAAGTACTGGAAGTGTGTGTC----- 1450
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1470 CTCACAGATTAATGCTGTAATTTCTCTGTCGCACAAATCGTCAAGATTTGATGTCAAAC 1529
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1451 -----AGGCAACATCTATGAGAGATCTTATCAACTCTCGGGAAGCTTT 1498
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1530 TCGTTCTGCTAAGACACAGGTATACAGCGGTGTTAAACGTTTTTGACAGAGGATTC 1589
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1499 GTAATGGTGATAAAAAATTTGAATTTCTTTCATTTTCAGAGCCAGTTGCGGAGATAAT 1558
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1590 CACATGTGTGGCAGGAGTACTCGTTTCTTTCATTTCTCATCTAACCCAGCTGAGGACAGG 1649
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1559 TCAGTGTGATGTTTGCATCAAGACCTGGCTTACTGCAATGATATAGAGCTTGGATG 1618
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1650 TCAGCATGGTTCTTCGACAGAGGACAGAACACACAGTGGAAACCATTTAGGAATGGATG 1709
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1619 GGTGATTTTCGCAGATCAAGAAATGCGCAAAATATGCTGCCAGACTTGGTCAATCTTTT 1678
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1710 GGGCGGTTAC---AAGTAAGATGAGCAAGCATGCGCTCGATGGGCGAGTGTTC 1766
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1679 GGTTCCTCCAGAGAGACTTTGAGTGTCTTTAGGCATGAGATTTGAAGTATTCGCCAIGTA 1738
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1767 TCCTCTACATATGCTACGGTGGTCTGCGACCCGCTGAGGTAAATGAGTCTCTTGATGAA 1826
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1739 AAGGTTTCATGAACACCTAGTCTTTTCTGATGGAATGTTGTAATAATATCTGCTGACTTT 1798
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1827 GTTGAACAT---AACGGGTACATTTTCTCTGATGGAATTTGCAAGATTACGTCGACCTT 1883
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1799 GCTCATAGAGTTGCTCAAAATGTTGGCTTCAATATA---CCCCATGCTGTTCCAGATT 1855
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1884 GCACCTGGAAGTGTCTCAGAGCTGCATTCACAGATTAATCCCCATCTGCTTACCAGATT 1943
QY 1856 CGTTATGGTGGATATAAAGGTGTGTGGGTGTGATCCGG-----ATTCAATCAATGAAG 1909
Db 1944 AGGTATGCGAGCTTCAAGGGTGTATATCTCTCTGCGAAGGAAAAATGATGGATACGA 2003
QY 1910 TTGTCTTTGAGAAAGACATGTCGAATATGAATACACACACATAAAGTTAGATGTCCTT 1969
Db 2004 CTTTCCCTGAGCGCGAGCATGCACAAGTTTGTAGTCTTAACCATATGTTGTAGAGTGTC 2063
QY 1970 GGATGGAGCAATATCAGCCCTTGTATCTTAATCGTCAACATGATTAAGCTCTCTGTAC 2029
Db 2084 TCGTGGACAAAGTTTACCGCAGGATCTTAATCGTCAGATTAATACATTAATCTCTCT 2123
QY 2030 CTTGGAGTGAAGATGAAGTCTTCGACAGAGCAAGCAAAAGGAGCTGTAGATCAGCTTGT 2089
Db 2124 TTGAATGCCCGGATGCTATCTTGTCTCAAAATGCGAGGAGCAATGTTATCTAATCTCA 2183
QY 2090 GCTATCTTGCATGATCTTTTGAAGGCACAGGAGGCTTTGGAATTCATGTCTCTCTGAG 2149
Db 2184 AATATTTTGCAGACTCTGATGTTGCTTTGA---CATTTGAACCGGCTCTTGTGCTGAG 2240
QY 2150 AACACTAATATCTCAAGGCAATGCTAAACHTGTGTTATAGCTGTGATGCTGAGCCCTTT 2209
Db 2241 CAAGGAACCACTGCAGCACTGATGTGTAGTGTGCTGTGCTATTCACCTGGAACCTGAG 2300
QY 2210 CTTTCAATGATGTTGCCAAACCTTCGCGCATCCAAAGTGTGCTGATTTGCGGACTAGATCA 2269
Db 2301 CTGAAGCAATGCTGTTAGCTATAGGTCTTCACAGGTGCTAGGTCTTTTGGAGAAGACA 2360
QY 2270 AGAATATTATCCAAATGGAAGCAATGATGGATGTTTGGATGAATCCAGAACCTTG 2329
Db 2361 AGGATTTTGTGCCCAAGGAGGTGTTGATGGGCTGCTGTGATGAATCTGGATGCTT 2420
QY 2330 GAATATGTCAGGTGTTGTTTTCAGTTTACTGTGCTGACATGAGAGTGTCTTGACGAT 2389
Db 2421 GAGCAAGGCACTGCTTATCCGGGCATCACTCCATCAC-TCAATATTTGCTGTGA 2479
QY 2390 TTACATCCATTTAATACAGAGATCCACCAACATATTTCAATCTCTGAAGGAATGCT 2449
Db 2480 GTATGATCAAGATTTGCTGTCAG---CAACACAAATGAGAGACCATTTCTGGTACTATC 2537
QY 2450 GTTGTGCAAAAAATCCATGCTGTGATCTGCTGTGATTTTAAAGCTGTAAT 2509
Db 2538 GTAAATGGCAAGATCCATGCTTCATCCAGGGATGCGGAATCTTGAAGCTGTTGAT 2597
QY 2510 GTTCGAGCGCTGCACACATGATGTTGTTATTTCCCTCAGAAAGGAAAAAGACT 2569
Db 2598 GTCCCTGAACATGATCACCTTGTGATGCTTGTGCTTCCCCAAGAAAGGTGAGAGGCG 2657
QY 2570 CATCCGAATGAATGTTCTGGGAGTATTTGGATGGGATATCTACTTTGCTTGGGAT 2629
Db 2658 CAGCGCAATGAAGCATCTGGGAGTATCTTGTATGGGATCTATCTTGTAAACATGGGAT 2717
QY 2630 CAAGACATGATCCCGC-----AAGGCAAGTCCAGCCGATGATATCTCCAGCACCC 2683
Db 2718 GAACCTTATACCACCTGGTAAAGAGTTTGAACCCCTATGACTACTCCCGAGCTGAA 2777
QY 2684 AGCATACAGTTGACCATGATGTCACATTTAGGAAGTTGAGAGTACTTACCAACTAT 2743
Db 2778 GCAAAACAACTGCCAGCGCAGTATCCCAACATGATATTTGTTGGTTCTTCTTGAAGAC 2837
QY 2744 ATTGTGATGACAGTTTGGGAATCAGCAATGCAATGCCATGCTGATTTTCAGACAGAGAA 2803
Db 2838 ATGTAATAGAGAACTGGGTCCAAATAGCAATGCTCATGTTTTCACGCTGATGAGC 2897
QY 2804 CCGATATGCCATGATGATCATCATCAAAAACTTCTGAGCTCTTTCAATTCAGTGTG 2863
Db 2898 GAGTATGAGCAATGATGAGAGTGTATTTCAGTTGGCAGAACTAGCAGCAACTGCTGTG 2957
QY 2864 GACTTTCCAAAGACTGCTGTGCTCCGCTGAATACCATCTCAGTTGCGCCTTAAGATAC 2923
Db 2958 GACTTCCCAAGACAGCAAAAAATGTTGTAATGCCAGCATCCCTTCGACCAAAATTTAT 3017

QY 2924 CCAGACTTCATGATAAGCCGAGACAGACCAGTATATCTCAGAAAGAGTATTGGAAG 2983
Db 3018 CTTGACTTCATGGGAAGGAGTATCTCTCTATAGATCAGAGAAGATCTTGAAGG 3077
QY 2984 CTTTTCAGGAAAGTGAAGGACAAACACCTCAGGCTAGCTCTATCCGACCTTCAACA 3043
Db 3078 CTTTATCGGTCTAATCAAG---AAGCCTCCAGCGATGATTTGGTTCCAGAAAGAACTTGC 3134
QY 3044 GATGTTCAAGGAGATCATATGATGCTGATATGGAAGTTGATGATTTGAAGATTACATT 3103
Db 3135 ACATCTAACAAATCGCTTATGATGAGATATGGAAGTTGCTGTCAGCTGATTTCTC 3194
QY 3104 GAGCAAGCTTTTGACTACAAACTGAATATGACAAACAGCTGGTATTAATTAATGGA 3163
Db 3195 TCGAGTGTCTGGCAGTGCAGTGTCTCATATGAAACACAACTGAACGCACTGCTCAAC 3254
QY 3164 TATGGCATATAAAACAGAGCTGA 3186
Db 3255 TATGGGTGCGCACTGAAGCAGA 3277

RESULT 7

AAH77704
ID AAH77704 standard; cDNA; 3591 BP.
XX
AC AAH77704;
DT 13-NOV-2001 (first entry)
XX
DE Nucleotide sequence of a plant SGS2 polypeptide.
XX
KW SGS2; RNA-dependent RNA polymerase; transgene silencing;
KW transgene stability; crop plant; viral resistance; ss.
XX
OS Arabidopsis thaliana.

Key Location/Qualifiers
CDS 1..3591
/*tag= a
/product= "SGS2"

FR2804128-A1.

27-JUL-2001.

26-JAN-2000; 2000FR-0001007.

26-JAN-2000; 2000FR-0001007.

(RHOB-) RHOBIO.
(INRG) INST NAT RECH AGRONOMIQUE.

Beclin C, Elmavan T, Mourrain P, Vaucheret H;

WPI; 2001-543303/61.

P-PSDB; AAG67235.

A new plant SGS2 gene involved in encoding an RNA-dependent RNA polymerase and in transgene silencing, increases transgene stability and expression in transgenic plants when it is inactivated -

Claim 4; Page 32-37; 46pp; French.

The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA-dependent RNA polymerase and is involved in transgene silencing. Inactivation of SGS2 is used to increase transgene stability and expression in plants, particularly crop plants, especially maize, corn, barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants. Overexpression of SGS2 can be used to increase resistance to viral infection in plants.

Sequence 3591 BP; 1022 A; 714 C; 858 G; 997 T; 0 other;

Query Match 8.28; Score 305.8; DB 22; Length 3591;
Best Local Similarity 51.08; Pred. No. 3.2e-73;
Matches 1013; Conservative 0; Mismatches 917; Indels 57; Gaps 10;

QY 1263 TAAGAGGGTCTCTAGTACACACCGTCAAGGTTATTTTGTGGTCCAGAGGTTAATGTTT 1322
DB 1232 TCAGNAGACTAGTATATACCCACACAGAGGCTATGCTTACCCACAGAGTTCAGCTCT 1291
QY 1323 CCAATCGGGTCTCCGCAATTTATCTGAAAGACATAGATAAATTTCTTCGGTTCCTTTT 1382
DB 1292 CCAACAGGGTACTCAGGAGATACAAAGCTGTGCTGAAAGATTTTTCGGGTACTTCA 1351
QY 1383 TTGATGAGGATGGGAGAAACGTATTC-----TACAGACTATTACCAAAA 1429
DB 1352 TGGATGAAGTATGGAGACCAATAATTCGAATGTCTCTCTTATTTGTTGCTCCTATTG 1411
QY 1430 GCAAGTACTGGAAGTGTGT-----CAGGACAAACATCTATGAGAGGATCTTAT 1478
DB 1412 TGAAGGATTGACATCAAGTCTTTCTCCAGAAAGACTAGCTTTTAAAGAGTGAAGA 1471
QY 1479 CAACCTCTCGGAAAGCTTTGTAATTTGGTGATAAAATTTGAATTTCTTGCATTTTCAT 1538
DB 1472 GCATTAATACCGATGGGTTTAACTATGTGTAGAAATACAGTTTCTTAGCATCTCAG 1531
QY 1539 CGAGCCAGTTGGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAA 1598
DB 1532 CCAATCAACTGAGAGCCGCTCTGCATGGTCTTTGCTGAGAGCGGAAACACGTGTGT 1591
QY 1599 ATGATATAGAGCTTGGATGGTGTATTTTCGAGATCAAGAAATGTCGCAAAATATGCTG 1658
DB 1592 CAGATATAAAACATGGAATGGGAGTTCAAAGA---CAAGAAATGGCAAAATGTGCTG 1648
QY 1659 CCAGACTTGGTCAATCTTTTGGTCTCCAGAGAGACTTTTCAAGTGTCTTTAGGATGAGA 1718
DB 1649 CTAGATGGGCTGTGCTTCTCTCCACATATGCCATGTAGATGTCATGCTCACGAGG 1708
QY 1719 TTGAAGTATATCCCATGTAAGGTTTCATGGAACAGCTATGCTTTCTGATGGAATG 1778
DB 1709 TTGA---CACTGAGTTCAGATATAGAGAAATGGGTATGTTTCTCTGACGGAATG 1765
QY 1779 GTAAATATCTGTGACATTTGCTATAGATGTGCTCAAAA-----TGTGGCCTTCAAT 1832
DB 1766 GTACATCACACCTCACCTCGCTGAGGAAGTAATGGAGAACTTAAGTTGATGCACT 1825
QY 1833 ATACCCATCTCTTCCAGATCTGTTATGGTGGATATAAGGTTGTTGGGTGTGATC 1892
DB 1826 ACAGCCCTTGTCTTATCAGATACGTTACGCAAGTTCCTCAAGGGGTGTTGCTCGTTGGC 1885
QY 1893 CGGATTCATC-----AATGAAGTTGTCTTTGAGAAAGAGCATGTCGAAATATGAATCAG 1946
DB 1886 CATCAAAAGTATGGAATCAGGCTAGCCCTTCGAGACAGTATGAAGAGTCTTTTCCA 1945
QY 1947 ACAACATAAAGTTAGATGTCCTTGGATGGAGCAAAATACAGCCTTGTATCTTAATGCTC 2006
DB 1946 AACATACGATCTTGAGATCTGTTCTGGACGAGGTTTCAACCTGGGTTCTTAAATCGGC 2005
QY 2007 AACTGATACGCTCTGTCTACATCTGAGTGAAGATGAAGTTCTCCAGCAGAGCAAA 2066
DB 2006 AGATAATTACCTCTTATCCGTACTAGGTGTTCCGGATGAATAATCTCGGATATGCAAG 2065
QY 2067 AGGAAGCTAGATCAGCTTGTATCTTCTGATCATCTTTTGAAGCAGCAGGAGGCTT 2126
DB 2066 AATCCATGCTCTATAAAGTGAACCGCATCTTGTATGATACAGATGTGGCATTTGAAGTTC 2125
QY 2127 TGGATTTGATGCTCTCTGAGAGACACTAATATCTCAAGCAATGCTAACTGTGGTT 2186
DB 2126 T---CACGGCATATGTGCTGAACAGGGAAACACTGCACTATCATGCTTATGTCAGGTT 2182
QY 2187 ATAAGCCTCATGTGAGCCCTTCTTCTCAATGATGTTGCAAACTTTCGCGCATCCAAAGT 2246
DB 2183 TCNAACCAAAACCGAGCGCATCTACGGGGATGTTGCTTCACTCAGATGTCAGCAAC 2242

RESULT 8
AAC63741
ID AAC63741 standard; cdna; 2816 BP.
XX

Db 836 GAAGCTGATATAGACCTGGATGGATGCTAGGGGCTCTATTGATGTTGTTTCCT 895
QY 2552 CAGAAAGGAAAGACCTATCCGAATGAATGTTCTGGAGTGTATTTGGATGGGATATC 2611
Db 896 CAGAGAGGGGAAAGGCTCATCCGAATGAATGCTCCGGGGCGATTTGGATGGGACCTC 955
QY 2612 TACTTTGTTGCTGGATCAAGACATGATCCCGCAAGCGAAGTCCAGCCGATGGAATAT 2671
Db 956 TCTTTTATTCTTGGATGACAACTGATTCGGGAGAAGTTGATGCACCTATGGATAC 1015
QY 2672 CUTCAGACCCAGCATACAGTGGACCATGATGATGTCACAAATGAGGAAGTTTGAAGATAC 2731
Db 1016 ACTGCAAGAGGCCACGCATATGACCATGCTGTACACTGAGGAATTCAGAAGCAC 1075
QY 2732 TTCACCACTATATGTAATGACAGTTTGGGAATCATAGCAATGCCCATGTCGTATT 2791
Db 1076 TTCGTGAGTACATGATAACGATACCTCGGTGCTCCAGCCGACCTTGATCCAC 1135
QY 2792 GCAGACAGAACCTGATATGCCATGAGTATGATGATCCATGCAAAAACCTTCTGCTGAGTCTTT 2851
Db 1136 GCAGACCTGATCCCTGAAAGCTCGACGCCGAGTGGTCCAGCTGGCGCTCGAC 1195
QY 2852 TCAATGTCAGTGGATTTCCAAAGACTGGTGTTCCTCCGCTGAAATACC 2898
Db 1196 TCATGGGGGTCGACTTCGCCAAGAGCGGAGCTCAGCCCAAGATCC 1242

RESULT 10

AAC63743

ID AAC63743 standard; cDNA; 740 BP.

AC AAC63743;

DT 07-FEB-2001 (first entry)

DE Maize RNA-directed RNA polymerase EST contig, SEQ ID NO: 13.

KW Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;

KW gene mapping; gene marker; plant virus resistance; plant breeding; ss.

XX Zea mays.

XX WO200060097-A1.

XX 12-OCT-2000.

XX 06-APR-2000; 2000WO-US09105.

XX 07-APR-1999; 99US-0128094.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX (PION-) PIONEER HI-BRED INT INC.

XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;

XX WPI: 2000-679276/66.

XX P-PSDB; AAE28534.

XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful

XX for controlling gene expression and providing mechanisms to engineer

XX plant virus resistance.

XX Disclosure; Page 55; 62pp; English.

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ Sequence 740 BP; 217 A; 141 C; 205 G; 166 T; 11 other;

Query Match 7.4%; Score 276.2; DB 21; Length 740;

Best Local Similarity 64.2%; Pred. No. 2e-65;

Matches 451; Conservative 0; Mismatches 231; Indels 21; Gaps 2;

QY 2771 GCAAAATCCCATCGCTATTGTCAGACAGAGAACCTTATATGGCCCATGAGTGCATGC 2830
Db 2 GCCAATCGCATGTGGTCTTTGTCAGATCAGGAACGATGAGGCTGAGAGTCCACCGTGC 61

QY 2831 AAAAACTTCTGAGCTCTTTTCAATGTCAGTGGAGCTTCCAAAGACTGTGTTCGGCT 2890
Db 62 GTTCAACTGGCCAGCTCTTCTATAGCTGTGATTTCCAAAGACTGTGAGTGCCTGCT 121

QY 2891 GAATATCCATCTGATGTCGCTTAAAGATACCCAGACTTCATGGATTAAGCCGACAA 2950
Db 122 CTGATTCACATGAGCTACATGTCNAGGAGTATCTGACCTCATGGAGAATCTGACAAA 181

QY 2951 ACCAGCTATATCTCAGAAAGAGTTATTGAAAGCTTTTCAGGAAAGTGAAGGACAA 3010
Db 182 GTCACCTATGAATCAAAGGGTGTGATCGGAAGCTCTATAGGAAATATAAGAAAGCAC 241

QY 3011 CCTCAGCTAGCTCTATTCGGGACCTTCACAAGAGATGTTGCAAGGAGATCATATGCT 3070
Db 242 CCACA-----CATAAAGCACTTCACGAGGAAGTGGCAAGCGGCTTTATGACACC 292

QY 3071 GATATGGAAGTTGATGATTGAAAGATTACATTGACCAAGCTTTTGACTACAAAAC 3130
Db 293 GATTGATTGTTGATGCTATGAGATTACATTACTGAGGCTATAGAGTTCAGGAAGAG 352

QY 3131 TATGACAAAGCTGGGTAAATTTAATGACTACTATGGCATATAAAGAGAGCTGAATA 3190
Db 353 TACGATTTTCAAGCTGGGTAAATTTAATGACTACTATGGCATATAAAGAGAGCTGA 412

QY 3191 CTTAGTGGTGGCATTTATGAAGGCTCAAAAACCTTTGACCGCAGAAAAGATGCTAGGCC 3250
Db 413 ATAAAGTGGATGTTCTAAAGATGGCAAGATTTTCACCAAGAGTAGNATGCTGATGCA 472

QY 3251 ATTAGTGTGTCAGGGCTTTGAGGAAGAGGCAAGAGCTGGTTCAGAGGCGTA-- 3308
Db 473 ATTAGAATGGCGNGAGATCTTTGAGGAAGAAGCTAGTTCGNGGNTCAATGAGATGAGC 532

QY 3309 -----ATGATATAGATGACATGTTACCAAAGCTTCGGCTTGGTACCAAGTTACA 3358
Db 533 ACAGGAGAGGATGGCCAGATGCCATGGAGGCCAAGGCTCTCTTGGNACCATGTTACT 592

QY 3359 TATCATCTACATATGGGTGCTACATCAGGGGTGAAAGAGCTCATTTTCATTAGC 3418
Db 593 TATCATCAGCAGTACTGGGGCAGCTACAATGAAGGATGATGATCGNGCCATCTTATTAGC 652

QY 3419 TTTCCCTGGTGTGTTTATGACCAAGCTTAATCCAGATTAAAGAGG 3461
Db 653 TTCCATGGNGCGGATATGACAAAGCTTGGGGGCGATCAAGCANG 695

RESULT 11

AAH77703

ID AAH77703 standard; DNA; 6863 BP.

AC AAH77703;

XX AAH77703;

DT 13-NOV-2001 (first entry)

DE Nucleotide sequence of the SGS2 gene.

KW SGS2; RNA-dependent RNA polymerase; transgene silencing;

KW transgene stability; crop plant; viral resistance; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

XX Promoter 1..850

XX /*tag= a

```

FT exon      851..3564
FT           /*tag= b
FT intron    3565..3986
FT           /*tag= c
FT exon      3987..4862
FT           /*tag= d
FT terminator 4863..6863
FT           /*tag= e
XX
XX FR2804128-A1.
XX
XX PD 27-JUL-2001.
XX
XX PF 26-JAN-2000; 2000FR-0001007.
XX
XX PR 26-JAN-2000; 2000FR-0001007.
XX
XX (RHOB-) RHOBIO.
XX (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX Beclin C, Elmayer T, Mourrain P, Vaucheret H;
XX WPI; 2001-543303/61.
XX
XX A new plant SGS2 gene involved in encoding an RNA-dependent RNA
XX polymerase and in transgene silencing, increases transgene stability
XX and expression in transgenic plants when it is inactivated -
XX
XX Claim 2; Page 30-32; 45pp; French.
XX
XX The present sequence encodes a plant SGS2 polypeptide. SGS2 is an
XX RNA-dependent RNA polymerase and is involved in transgene silencing.
XX Inactivation of SGS2 is used to increase transgene stability and
XX expression in plants, particularly crop plants, especially maize, corn,
XX barley, sorghum, soy, sugar cane, beet, tobacco or cotton plants.
XX Overexpression of SGS2 can be used to increase resistance to viral
XX infection in plants.
XX
XX Sequence 6863 BP; 1862 A; 1354 C; 1485 G; 2162 T; 0 other;
XX
XX Query Match      5.9%; Score 220.4; DB 22; Length 6863;
XX Best Local Similarity 51.5%; Pred. No. 1.4e-49;
XX Matches 733; Conservative 0; Mismatches 641; Indels 48; Gaps 8;
XX
XX QY 1263 TAAGAGGGTCTAGTACACCATCAAGTTTATTTTGGTCCAGAGGTTAATGTTT 1322
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2081 TCAGAAGACTAGTAATACCCCAACAGAGCTATGCTACCCCAAGAGTTAGCTCT 2140
XX
XX QY 1323 CCAATCGGTTCTCGGCAATTATCTGAAGACATAGATAAATTTCTGCTTTCTTTG 1382
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2141 CCAACAGGCTACTCAGGAGATCAAAAGCTGTGCTGAAAGATTTTGGGGTAACTTCA 2200
XX
XX QY 1383 TTGATGAGAGTGGGAGAACTGTATTC-----TACAGACTTATTACCAAAA 1429
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2201 TGGATGAAAGTATGAGACCAATAAATCGAATGTTCTCTCTACTTTGTTCTCTATG 2260
XX
XX QY 1430 GCAAGTACTGGAAGTGGT-----CAGGACAAACATCTATGAGAGGATCTTAT 1478
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2261 TGAAGGATTGACATCAAGTTCTTCTCCAGAGACCTACGTTTAAAGAGTGAAGA 2320
XX
XX QY 1479 CAACCTCGCGGAAAGCTTTGTAATGTTGATATAAAATTTGAATTTCTTCATTTTCAT 1538
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2321 GCATATTACCGATGGGTTAACTATGTTGTAAGAAATACAGTTTCTTAGCATTTCTAG 2380
XX
XX QY 1539 CGAGCCAGTTGGGATTAATTCAGTGTGATGTTTGCATCAAGACCTGGCCTTACTGCAA 1598
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2381 CCAATCACTGAGACCCGCTCGATGGTCTTTCGTAAGACCGGAAACACGTGTGT 2440
XX
XX QY 1599 ATGATATAAGACTTGGATGGGTGATTTTTCGAGATCAAGAAATGCCAAATAATGCTG 1658
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 12
ABL71773

```

DB 2441 CAGATATAAAACATGGATGGGGAAGTT---CAAAGACAAAGAAATGTGGCAAATGTGCTG 2497
QY CCAGACTTGGTCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTAGCAGTACGAGA 1718
DB 2498 CTAGGATGGCCCTGTGCTTCTCTCCACATATGCCACTGTAGATGTCATGCCCTCAGAGG 2557
QY TTGAAGTTATCCCGATGTAAAGGTTTCATGGAAACAGCTATGTCTTTCTGTGATGGAATTG 1778
DB 2558 TTGA---CACTGAGGTTCCAGATATTGAGAGAAATGGGTATGTTTCTCTGACGGAATTG 2614
QY GTAAATATCTGCTGACTTTGCTCATAGAGTTGCCCTCAAAA-----TGTGGCTTCAAT 1832
DB 2615 GTACATCACACCTGACCTGCTGACGAGTAATGAGAACTTAAGTTGGATGTCACCT 2674
QY ATACCCCATCTGCTTCCAGATTGCTTATGGTGGATATAAAGCTGTGTGGGTGTTGATC 1892
DB 2675 ACAGCCCTTGTGCTTATCAGATACGTTACGAGGTTTCAAAGGGGTTGTGCTCGTTGGC 2734
QY C-----GGATTTCATCAATGAAGTTGCTTTGAGAAGAGCATGTCGAAATATGAATCAG 1946
DB 2735 CATCAAAAAGTGATGGAATCAGGCTAGCCCTTCGACAGATGACAGAGATCTTTTCCA 2794
QY ACAACATAAAGTTAGATGTCCTTGGATGGAGCAAAATATCAGCCTTGTATCTTAAATCGTC 2006
DB 2795 AACATACGATCTTGGAGATCTGTTCTCTGGACGAGGTTTCAACCTGGGTTCTTAAATCGC 2854
QY AACTGATTACGCTCTGTTCTACACTTGGAGTGAAGTGAAGTCTCGAACAGCAACAA 2066
DB 2855 AGATAATTACCTTCATCCGTTACTAGGTTTCCGGATGAAATATTTCTGGATATCGAG 2914
QY AGGAAGCTGTAGATCAGCTTGTATCTTGTGATGATTCTTTTGAAGGCACAGAGGCTT 2126
DB 2915 AATCCATGCTCTATAAACTGAACCGCATCTTGATGATACAGATGTGGCATTTGAAGTTC 2974
QY TGGAAATTGATGCTCTCTGGAGAGAACACTAATATTCTCAGGCAATGCTTAACTGTTG 2186
DB 2975 TCACGGCATCATGTGCTGTAACAGGAAACACTGCACGTAT---CATGCTTAGTGCAGTT 3031
QY ATAAGCCTGTATGTAGCCCTTTCTTTTCAATGATGTGTTGAAACCTTCCGCGCATCAAAGT 2246
DB 3032 TCAACCAAAAACCGAGCCGCACTACGCGGGATGTTGTTCTTCAGTCAGAATTCACAAC 3091
QY TGCTCGATTTGGGACTAGATCAAGATATTTTATTCAAATGGAAGAACAAATCATGGAT 2306
DB 3092 TCTGGGCTCTCAGAGAAAATCTCGTATTTTGTACTTCAGGAAGGTGGCTAATGGTT 3151
QY GTTTGGATGAATCCAGAACTTGGATATGTTGTCAGGTGTTTGTTCAGTTTACTTGTGCTG 2366
DB 3152 GCCTAGACGAGCAGGATACTTGAACATGSCCAATGCTTATTCAAGTCTCTAAACCGT 3211
QY GACATGAGAGTGTCTGAGATTTACATCCATTTAATACACAGCATCCACCACAGTA 2426
DB 3212 CTATAGAAAATTTGTTCTCCAAACATGGGTCTCGTTTAAAGGAGACAAAGACAGATCTGG 3271
QY ATTTTCATCTGAAGGAAATGTGTTGTTGCAAAAATCCATCTTCATCTCTGTTGATA 2486
DB 3272 AGTAGTT---AAAGCTATGAGCATGCTTAAGATCCTTCTCTCACCAGGGGATG 3328
QY TTCGTGTTTAAAGCTGTAAATGTTTCGAGCGCTGCACACAGCTGATGTTGTTGAT 2546
DB 3329 TAAGGATTTTAGAAGCTGTTGATGTATCCCGAGCTGCATCATGATGATGCTGCTTATTT 3388
QY TCCTCAGAAAGGAAAGAACCTCATCCGAATGAATGTTCTCTGGAGTGATTTGGATGGGG 2606
DB 3389 TCCCTCAGAAAGGTATAGCCCGCATACAACGAGCTTCTGCGAGTGACCTTGACGGG 3448
QY ATATCTACTTTTGTGCTGGATCAAGACATGATCCGGCAA 2648
DB 3449 ACCTGTACTTTGTGGCTTGGGATCAGAAACTCATCCCTCCA 3490

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 19:22:12 : Search time 158 seconds
(without alignments)
7867.093 Million cell updates/sec

Title: US-09-782-874-1
Perfect score: 3731
Sequence: 1 GAAATATCTTACTTACTT.....AGTTTCATCTTCTCTCAA 3731

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PC1_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PC1US_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3731	100.0	3731	10	US-09-782-874-1
2	106.8	2.9	282	10	US-09-294-093B-1147
3	95.4	2.6	274	10	US-09-294-093B-1096
4	75	2.0	260	10	US-09-878-574-10527
5	40.4	1.1	23071	10	US-09-764-864-1673
6	40.2	1.1	1904	10	US-09-815-242-8362
7	39.2	1.1	2073	10	US-09-815-242-4555
8	39.2	1.1	2109	10	US-09-815-242-4555
9	38	1.0	5191	10	US-09-962-832-124
10	37.8	1.0	2720	10	US-09-847-665-3
11	37.8	1.0	3699	10	US-09-847-665-6
12	37.6	1.0	567	10	US-09-974-300-1446
13	36.4	1.0	499	10	US-09-864-761-412
14	36.4	1.0	3191	10	US-09-939-980-194
15	36.2	1.0	1196	10	US-09-764-847-1401
16	36	1.0	273	10	US-09-923-876-6245
17	35.6	1.0	2397	10	US-09-909-320-324
18	35.6	1.0	2397	10	US-09-909-088B-324
19	35.6	1.0	2397	12	US-10-052-586-29

C	20	35.4	0.9	53226	10	US-09-818-264-3	Sequence 3, Appli
	21	35	0.9	1275	10	US-09-771-838A-4	Sequence 4, Appli
	22	35	0.9	1458	10	US-09-771-838A-19	Sequence 17, Appl
	23	35	0.9	1796	10	US-09-771-838A-17	Sequence 19, Appl
	24	35	0.9	1845	10	US-09-771-838A-5	Sequence 5, Appli
C	25	35	0.9	2000	10	US-09-887-575-126	Sequence 126, App
	26	35	0.9	2046	10	US-09-771-838A-6	Sequence 6, Appli
	27	35	0.9	3023	10	US-09-079-892-4	Sequence 4, Appli
	28	35	0.9	3048	12	US-10-044-090-685	Sequence 685, App
	29	35	0.9	111282	12	US-10-094-989-3	Sequence 3, Appli
	30	35	0.9	397658	10	US-09-813-320-3	Sequence 3, Appli
	31	34.6	0.9	471	10	US-09-815-242-4575	Sequence 4575, Ap
	32	34.6	0.9	549	10	US-09-815-242-8154	Sequence 8154, Ap
C	33	34.6	0.9	1134	10	US-09-815-242-7451	Sequence 7451, Ap
	34	34.6	0.9	1883	10	US-09-796-256A-3	Sequence 3, Appli
	35	34.6	0.9	4756	10	US-09-982-091A-3	Sequence 3, Appli
C	36	34.6	0.9	32146	10	US-09-764-860-797	Sequence 797, App
C	37	34.6	0.9	32248	10	US-09-764-860-802	Sequence 802, App
	38	34.4	0.9	304	10	US-09-294-093B-623	Sequence 623, App
C	39	34.4	0.9	329	10	US-09-960-352-7689	Sequence 7689, Ap
	40	34.4	0.9	501	10	US-09-864-761-12414	Sequence 12414, A
	41	34.4	0.9	1537	10	US-09-740-273-1	Sequence 1, Appli
C	42	34.4	0.9	8188	10	US-09-070-927A-21	Sequence 21, Appli
	43	34.2	0.9	3831	12	US-10-051-952-5	Sequence 5, Appli
C	44	34	0.9	42999	10	US-09-740-028-3	Sequence 3, Appli
C	45	33.8	0.9	1867	10	US-09-846-590B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-782-874-1
; Sequence 1, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassenegeger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDMP)

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/09/782,874

APPLICATION NUMBER: US/09/782,874

FILING DATE: 08-Feb-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/811,583

FILING DATE: 05-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Haley, James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: MPG-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-596-9000

TELEFAX: 212-596-9090

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3731 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Tomato
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 194..3535
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-782-874-1

Query Match      100.0%; Score 3731; DB 10; Length 3731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATATCTTTACCTTACTTCCACAGGAGTTCAGCTATCATCTCCCTCAAGTCTTTGGT 60
DB 1 GAAATATCTTTACTTTACTTCCACAGGAGTTCAGCTATCATCTCCCTCAAGTCTTTGGT 60
QY 61 GTTGTCATAATAAATTTGGTGTGCTTCAGTTTCAGTTCAGTCTGCTGGGTAGTTTAT 120
DB 61 GTTGTCATAATAAATTTGGTGTGCTTCAGTTTCAGTTCAGTCTGCTGGGTAGTTTAT 120
QY 121 TTTGTCATAACTTCAGGGGGTATTCAGTTGGTGTAGCATTTGAAAGTCGAAGTGCACCT 180
DB 121 TTTGTCATAACTTCAGGGGGTATTCAGTTGGTGTAGCATTTGAAAGTCGAAGTGCACCT 180
QY 181 GGAATTTGGCTACATGGGAAGACAAATTCAGGTTTTCGGATTCCCTTATCTTCTCTGC 240
DB 181 GGAATTTGGCTACATGGGAAGACAAATTCAGGTTTTCGGATTCCCTTATCTTCTCTGC 240
QY 241 GGAAGTGGTAAAGTCAATCTTTAGAGAAATATACAGGATATGGAAGTATGATGCAATGGA 300
DB 241 GGAAGTGGTAAAGTCAATCTTTAGAGAAATATACAGGATATGGAAGTATGATGCAATGGA 300
QY 301 GGTAAACAGTCCCAAGAGGATCTAGAGCATTTGCCAAAGTTCAATTTGCCGCAACAT 360
DB 301 GGTAAACAGTCCCAAGAGGATCTAGAGCATTTGCCAAAGTTCAATTTGCCGCAACAT 360
QY 361 AAGTGTGTGACAAATCATCACTTTGGCTAAATACAGGCTGTATTTGGCTCTTCTTATTT 420
DB 361 AAGTGTGTGACAAATCATCACTTTGGCTAAATACAGGCTGTATTTGGCTCTTCTTATTT 420
QY 421 GAAGGCTTGGGAAATGAACAACTGATTTGTCCTCAACTGGGCAATATGTTGGATCAGATGGA 480
DB 421 GAAGGCTTGGGAAATGAACAACTGATTTGTCCTCAACTGGGCAATATGTTGGATCAGATGGA 480
QY 481 TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATGACAAAGTTTGCAGTGTGGGGAAG 540
DB 481 TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATGACAAAGTTTGCAGTGTGGGGAAG 540
QY 541 TACAGAAGTTTCAATTCGAATTTGGCATTTGGATTGAAGAAATTTTTCCTTTTATCTAG 600
DB 541 TACAGAAGTTTCAATTCGAATTTGGCATTTGGATTGAAGAAATTTTTCCTTTTATCTAG 600
QY 601 TGGTTCAGCTGACTATAAATTCAGCTTTCATATGAAATATATGCGCAGTTTGTCTCCA 660
DB 601 TGGTTCAGCTGACTATAAATTCAGCTTTCATATGAAATATATGCGCAGTTTGTCTCCA 660
QY 661 TCGTCCATATGTCGCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTGTCTCCGAT 720
DB 661 TCGTCCATATGTCGCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTGTCTCCGAT 720
QY 721 CTATAGAGACATTGAAACTCCTGTATAGCTTCTTTAAGGAACTCCTGATGATCAGTG 780
DB 721 CTATAGAGACATTGAAACTCCTGTATAGCTTCTTTAAGGAACTCCTGATGATCAGTG 780
QY 781 GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCTAGCTTATGTTTCCA 840
DB 781 GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCTAGCTTATGTTTCCA 840
QY 841 GTTCCGTAGGGGGTTCGCTCTTCCAAATTTCCAGGAAAGTTTTCCTACTATGCGAGAACG 900
```

```

DB 841 GTTCCGTAGGGGGTTCGCTCTTCCAAATTTCCAGGAAAGTTTTCCTACTATGCGAACG 900
QY 901 TGAACCAATATTACTTTTACAGACTGGTTTCACTTTTTCGTCTCTCAAAATTCGGCTCT 960
DB 901 TGAACCAATATTACTTTTACAGACTGGTTTCACTTTTTCGTCTCTCAAAATTCGGCTCT 960
QY 961 GGTTCCTCAATGTCAGCTCCGGAAGGAATTTCAATTCCTTCAAGATTTTGTTCAAAAT 1020
DB 961 GGTTCCTCAATGTCAGCTCCGGAAGGAATTTCAATTCCTTCAAGATTTTGTTCAAAAT 1020
QY 1021 TAGTTCCTTTGGTACAGCATGGATGATACCTGGGCGAGCATTTAAATGCTACTTTTCCG 1080
DB 1021 TAGTTCCTTTGGTACAGCATGGATGATACCTGGGCGAGCATTTAAATGCTACTTTTCCG 1080
QY 1081 ATTAGTTTGTCTCCGAGGAGAAATGCGCATGCAATGAGCATGCCCTTAGAGAACTGTA 1140
DB 1081 ATTAGTTTGTCTCCGAGGAGAAATGCGCATGCAATGAGCATGCCCTTAGAGAACTGTA 1140
QY 1141 CTATATAAAGAGTGTCTGTTTATGATCCGTTGAGTGGCTCACTGAGCAGTATGATGGTA 1200
DB 1141 CTATATAAAGAGTGTCTGTTTATGATCCGTTGAGTGGCTCACTGAGCAGTATGATGGTA 1200
QY 1201 TCTCAAGGTTAGACAACCTCCAAATCTCCGTCATCCACTTTAGATGATGGTGTGTA 1260
DB 1201 TCTCAAGGTTAGACAACCTCCAAATCTCCGTCATCCACTTTAGATGATGGTGTGTA 1260
QY 1261 TGTAAAGAGGTTCTTAGTAACACCATGCAAAAGTTTATTTTGTGGTCCAGAGTTAATGT 1320
DB 1261 TGTAAAGAGGTTCTTAGTAACACCATGCAAAAGTTTATTTTGTGGTCCAGAGTTAATGT 1320
QY 1321 TTCCAATCGGGTCTCCGCAATTTCTGAGACATAGATAAATTTCTTCGTCTTCTTT 1380
DB 1321 TTCCAATCGGGTCTCCGCAATTTCTGAGACATAGATAAATTTCTTCGTCTTCTTT 1380
QY 1381 TGTGTGATGAGAGTGGGAGAACTGTATTTCTACAGACTTATACCAAAAGCAAGTACTGG 1440
DB 1381 TGTGTGATGAGAGTGGGAGAACTGTATTTCTACAGACTTATACCAAAAGCAAGTACTGG 1440
QY 1441 AAGTGGTTCAGGACAAACATCTATGAGAGATCTTATCAACTTCGGGAAGGCTTGT 1500
DB 1441 AAGTGGTTCAGGACAAACATCTATGAGAGATCTTATCAACTTCGGGAAGGCTTGT 1500
QY 1501 AATTTGTGATAAAAAATTTGAAATTTCTTGCATTTTCATCGAGCCAGTTGCGGATAATTC 1560
DB 1501 AATTTGTGATAAAAAATTTGAAATTTCTTGCATTTTCATCGAGCCAGTTGCGGATAATTC 1560
QY 1561 AGTGTGATGTTTGCATCAAGACCTGGCCTTACTGCAAAATGATATAGACTTTGGATGGG 1620
DB 1561 AGTGTGATGTTTGCATCAAGACCTGGCCTTACTGCAAAATGATATAGACTTTGGATGGG 1620
QY 1621 TGAATTTTCGAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680
DB 1621 TGAATTTTCGAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680
QY 1681 TTCTCCAGAGAGACTTTGAGTGTCTTAGGCATGAGATTGAAGTTATTTCCCATGTAAA 1740
DB 1681 TTCTCCAGAGAGACTTTGAGTGTCTTAGGCATGAGATTGAAGTTATTTCCCATGTAAA 1740
QY 1741 GGTTCATGGAACCAAGCTATCTTCTGATGGAATTTGTAATAATCTCGTCACTTTCG 1800
DB 1741 GGTTCATGGAACCAAGCTATCTTCTGATGGAATTTGTAATAATCTCGTCACTTTCG 1800
QY 1801 TCATAGAGTTTGCCTCAAAATGTCGCCCTTCAATATACCCCATCTGCTTCCAGATTCGTTA 1860
DB 1801 TCATAGAGTTTGCCTCAAAATGTCGCCCTTCAATATACCCCATCTGCTTCCAGATTCGTTA 1860
QY 1861 TGGTGGATATAAAGGCTGTTGGGTTGATCCGATTCATCAATGAAGTTGCTTTGAG 1920
DB 1861 TGGTGGATATAAAGGCTGTTGGGTTGATCCGATTCATCAATGAAGTTGCTTTGAG 1920
QY 1921 AAAGAGCATGTCCGAAATATGAATCAGACAAACATAAAGTTAGATGCTCCTTGGATGGAGCAA 1980
```


Db	1921	AAAGAGCATGCGAAATATGAATCAGACAACATAAAGTTAGATGTCCCTTGGATGGAGCAA	1981
Qy	1981	ATATCAGCCCTGTTATCTTAAATCGTCAACTGATTACGCTCTGTCTACACTTTGGAGTAA	2040
Db	1981	ATATCAGCCCTGTTATCTTAAATCGTCAACTGATTACGCTCTGTCTACACTTTGGAGTAA	2040
Qy	2041	AGATGAAGTCTCGAACAGAGCAAAAGGAAGCTGTAGATCAGCTTGAATGCTATCTTGGCA	2100
Db	2041	AGATGAAGTCTCGAACAGAGCAAAAGGAAGCTGTAGATCAGCTTGAATGCTATCTTGGCA	2100
Qy	2101	TGATTCCTTTGAAGGCACAGAGGCTTTGGAATTGATGTCTCCTCGAGAGAACACTAATAT	2160
Db	2101	TGATTCCTTTGAAGGCACAGAGGCTTTGGAATTGATGTCTCCTCGAGAGAACACTAATAT	2160
Qy	2161	TCTCAGGCAATGCTTAACTGTGGTTATTAAGCCTGATGCTGAGCCCTTTCTTCCAATGAT	2220
Db	2161	TCTCAGGCAATGCTTAACTGTGGTTATTAAGCCTGATGCTGAGCCCTTTCTTCCAATGAT	2220
Qy	2221	GTTTGCAAACCTTCGCGGCATCCAAAGTTGCTCGATTGGGACTAGATCAAGAATATTAT	2280
Db	2221	GTTTGCAAACCTTCGCGGCATCCAAAGTTGCTCGATTGGGACTAGATCAAGAATATTAT	2280
Qy	2281	TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAAATATGTCA	2340
Db	2281	TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAAATATGTCA	2340
Qy	2341	GGTGTGTTGTCAGTTTACTTACTTGGCTGGACATGGAGAGTTTCTGACGATTTACATCCATT	2400
Db	2341	GGTGTGTTGTCAGTTTACTTACTTGGCTGGACATGGAGAGTTTCTGACGATTTACATCCATT	2400
Qy	2401	TATATACAGCAGATCCACCACAGTAATTTCAATCTGAAGGGAATGTGGTTGTGCAAA	2460
Db	2401	TATATACAGCAGATCCACCACAGTAATTTCAATCTGAAGGGAATGTGGTTGTGCAAA	2460
Qy	2461	AAATCCATGCTTCATCCTCGTGATATTGCTGTTTTAAAGGCTGTAAATGTTCCGAGCGCT	2520
Db	2461	AAATCCATGCTTCATCCTCGTGATATTGCTGTTTTAAAGGCTGTAAATGTTCCGAGCGCT	2520
Qy	2521	GCACCACATGTTAGATTGTTGTATTCCCTCAGAAAGGAAAAAGACCTCATCCGAATGA	2580
Db	2521	GCACCACATGTTAGATTGTTGTATTCCCTCAGAAAGGAAAAAGACCTCATCCGAATGA	2580
Qy	2581	ATGTTCTGGAGTGATTTGGATGGGGATATCTACTTTGTTGCTGGGATCAAGACATGAT	2640
Db	2581	ATGTTCTGGAGTGATTTGGATGGGGATATCTACTTTGTTGCTGGGATCAAGACATGAT	2640
Qy	2641	CCGCGCAAGCAGTCCAGCCGATGGAATATCCTCCAGCACCAGACATCAGTTGGACCA	2700
Db	2641	CCGCGCAAGCAGTCCAGCCGATGGAATATCCTCCAGCACCAGACATCAGTTGGACCA	2700
Qy	2701	TGATGTCACAATTGAGGAAGTTGAAGATGACTTACCACATATATTGTGAATGACAGTTT	2760
Db	2701	TGATGTCACAATTGAGGAAGTTGAAGATGACTTACCACATATATTGTGAATGACAGTTT	2760
Qy	2761	GGGAATCATAGCAAAATGCCCATGCTGTAATTTCCAGACAGAACCTCATATGCCCATGAG	2820
Db	2761	GGGAATCATAGCAAAATGCCCATGCTGTAATTTCCAGACAGAACCTCATATGCCCATGAG	2820
Qy	2821	TGATCCATGCAAAAAACTTGCCTGAGCTCTTTTCAATTGCAGTGGACTTTCCAAAGACTGG	2880
Db	2821	TGATCCATGCAAAAAACTTGCCTGAGCTCTTTTCAATTGCAGTGGACTTTCCAAAGACTGG	2880
Qy	2881	TGTTCCGCTGAAATACCATCTCAGTTGGCCCTTAAGAATACCCAGACTTCATGGATAA	2940
Db	2881	TGTTCCGCTGAAATACCATCTCAGTTGGCCCTTAAGAATACCCAGACTTCATGGATAA	2940
Qy	2941	GCGGGAACAAGACCACTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGA	3000
Db	2941	GCGGGAACAAGACCACTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGA	3000
Qy	3001	GGACAAAGCACTTCAGGCTAGCTCTATCGGACCTTCACAGAGATGTTGCAAGGAGATC	3060
Db	3001	GGACAAAGCACTTCAGGCTAGCTCTATCGGACCTTCACAGAGATGTTGCAAGGAGATC	3060

RESULT 2

```

RESOLV 2
US-09-294-093B-1147
: Sequence 1147, Application US/09294093B
: Patent NO. US20010051335A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Lalgudi, Raghunath, V.
:
: APPLICANT: Ito, Laura, Y.
:
: APPLICANT: Sherman, Bradley, K.
:
: TITLE OF INVENTION: POLYNUCLEOTIDES AND P
:
: FILE REFERENCE: PL-0009 US
:
: CURRENT APPLICATION NUMBER: US/09/294,093B
:
: CURRENT FILING DATE: 1999-04-16
:
: PRIOR APPLICATION NUMBER: 60/082,567
:
: PRIOR FILING DATE: April 21, 1998
:
: NUMBER OF SEQ ID NOS: 6207
:
: SOFTWARE: PERL Program
:
: SEQ ID NO 1147
:
: LENGTH: 282
:
: TYPE: DNA
:
: ORGANISM: Zea mays
:
: FEATURE:
:
: NAME/KEY: misc_feature
:
: OTHER INFORMATION: Incyte ID No. US200100

```

US-09-294-093B-1147

Query Match 2.9%; Score 106.8; DB 10; Length 282;
Best Local Similarity 68.7%; Pred. No. 2.1e-19;
Matches 147; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 2440 GGGAAATGTGTTGTCACAAATCATCTTGCATCCTCGTGTGATATTCGTGTTTAAA 2499
DB 14 GGGTACTATCGTAATGCAAGAAATCCATCCCTTCATCAGGGGATGCGGAATCCTGA 73

QY 2500 GCGTGTAAATGTTGCGAGCGGTGCACACATGTTAGATTTGTGTATTCCTCAGAAAG 2559
DB 74 AGCTGTGTATGTCCTGAACTGACATCACCTTGTGATTTGCTGCTTCCCAAGAAAG 133

QY 2560 AAAAAGACCCATCCCGAATGAATGTTCTGGAGTGTATTTGGATGGGATATCTACTTGT 2619
DB 134 TGAGAGCCCCACGCGAATGAACATCTGGAGTGTATTTGATGGGATCTATCTACTTCT 193

QY 2620 TTGCTGGGATCAGACATGATCCCGCAAGGCA 2653
DB 194 AACATGGGATGAACACCTTATCCACCTGTGTAA 227

RESULT 3

US-09-294-093B-1096
; Sequence 1096, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Ialguidi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294, 093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 1096
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700343770H1
US-09-294-093B-1096

Query Match 2.6%; Score 95.4; DB 10; Length 274;
Best Local Similarity 65.7%; Pred. No. 2.6e-16;
Matches 157; Conservative 0; Mismatches 76; Indels 6; Gaps 1;

QY 2447 GTGGTTGTCGAAATCCATGCTTCATCCTCGTGTGATATTCGTGTTTAAAGGCTGA 2506
DB 2 GTTGCAATCAGAAACCCCTTGTCTCATCTCGGTGATGTCAGATCTTGAAGCTGTA 61

QY 2507 AATGTTGAGCGGTGCACCATAGG-----TAGATGTGTTGTTATTCCTCAGAAAGGA 2560
DB 62 TATGACCCCTGGATTGGATGCTAGGGGCTTATTGATTGTTGTTATTCCTCAGAGAGG 121

QY 2561 AAAAGACCTCATCGAATGAATGTTCTGGAGTGTATTTGATGGGATATCTACTTGT 2620
DB 122 GAAAGGCTCATCGAATGAATGCTCGGGGGGATTTGGATGGCAGCTCTTCTTTATT 181

QY 2621 TGCTGGATCAAGACATGATCCCGCAAGGCAAGTCCAGCGATGGAATATCCCTCAGC 2679
DB 182 ACITGGATGACAACTGATTCGCGAGAGGTTGATGACCTATGGACATACATGCAAC 240

RESULT 4

US-09-878-574-10527
; Sequence 10527, Application US/09878574
; Patent No. US20020110548A1

; GENERAL INFORMATION:

; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 10527
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 7009566775H1
US-09-878-574-10527

Query Match 2.0%; Score 75; DB 10; Length 260;
Best Local Similarity 64.7%; Pred. No. 8.8e-11;
Matches 130; Conservative 0; Mismatches 65; Indels 6; Gaps 1;

QY 2440 GGGAAATGTGTTGTCACAAATCCATGCTTCATCCTCGTGTGATATTCGTGTTTAAA 2499
DB 58 GGGAAAGTGTGGTGTCACAAATCCCTTCACCCAGAGACATCAGATCCTTGA 117

QY 2500 GGCTGTAATGTTGAGCGCTGCACCATGGT-----AGATTGTTGTATTCCCTCA 2553
DB 118 TGCTATCTACAGTGAAGATTAGAGGAAGGTTTGAGGATTCCTTGTATTCCACA 177

QY 2554 GAAAGAAAAAGACCTCATCCGAATGAATGTTCTGGAGTGTATTTGGATGGGATATCTA 2613
DB 178 AAAGACATAGGCTCATCCAAATGAATGCTCTGAGAGTATCTCGATGAGATTGTT 237

QY 2614 CTTTGTCTGGGATCAAGA 2634
DB 238 TTTCATAAGTTGGACAAAGA 258

RESULT 5

US-09-764-864-1673/c
; Sequence 1673, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1673
; LENGTH: 23071
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-1673

Query Match 1.1%; Score 40.4; DB 10; Length 23071;
Best Local Similarity 50.0%; Pred. No. 3.5;
Matches 101; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 2030 CTTGAGTGAAGATGAAGTCTCGAACAGAGCAAGAAAGAGCTGTAGATCAGCTTGT 2089
DB 12771 CTTGATTTCAAACCTGTGTAGTTGATGGCAAAATTAGGAACCAAGACTCAGAAATTA 12712

QY 2090 GCTATCTTCATGATCTTTGAAGGCACAGAGGCTTTGGAAATTCATGCTCTCGGAG 2149
DB 12711 AGTGTGTATAGATCTTGGAGGAAATTTATCTAGATTAATGCTCTCTGAGTCT 12652

QY 2150 AACACTAATATTTCTCAAGSCAATGCTAAACTGTGTTTAAAGCCTGATCGTGAAGCCCTTT 2209

	Query Match	1.1%	Score 39.2	DB 10	Length 2109
	Best Local Similarity	45.7%	Pred. No. 1.7		
	Matches 137	Conservative 0	Mismatches 163	Indels 0	Gaps 0
QY	2965	AGAAAGATTATTGGAAGCTTTTCAGGAAGTGAAGGACAAAGCACCTCAGGCTAGCTC	3024		
Db	453	AAAGAAGCTGTTGACGATTTTAAATAGTGAAGTAAAGAAATTCATATAAGAAAAAGA	512		
QY	3025	TATCGGACCTTTCACAAGAGATGTTTGAAGGAGATCATATGATGCTGATATGGAAGTTGA	3084		

	Query Match	1.0%;	Score 38;	DB 10;	Length 5191;
	Best Local Similarity	47.5%;	Pred. No. 6.4;		
	Matches 113;	Conservative	0;	Mismatches 125;	Indels 0; Gaps
Qy	257	TTCTTAGAGAAATATACAGGATATGGAACGTGATGTCATGGAGTTTAAACAGTCCAAA	316		
Db	1880	TACCTGGTAAATATTAAGCAGAAATCTCTATTGATCAACCTTATATTTCCTAA	1821		
Qy	317	GGAGATCTTAGAGCATTTGCCAAGTTCAAATTTGGCGGACACATAAAGTGCTGCAAAATC	376		
Db	1820	GTGTTATATATAGTTTTTCTTAAATCTTCAATTTTAAACATTAATTGACTGCTAAGATATATA	1761		
Qy	377	ATCACTTTGGCTAATACAGCGTGTTATTTGGCTCTCTTATTTCAAGCGCTGGGAATG	436		
Db	1760	TTCTGTTTTCGATAAAGAGATGTTTGTTTACCAAGCATCAAGAAATAGCTTTGCTATCT	1701		
Qy	437	AAAACATGATATTGCCAACTGCGGGCATATGTGGATCAGATGGATGGCATAACTTTGA	494		
Db	1700	AAGGATCACATTTAGACATAGGAAACAGCGTGTAGGTTCAGAAAGATGTGCTTACCTTCA	1643		

RESULT 10
US-09-847-665-3
; Sequence 3, Application US/09847665
; Patent No. US20020102566A1
; GENERAL INFORMATION:
; APPLICANT: ROGNER, UTE
; APPLICANT: SPYROPOULOS, DEMETRI
; APPLICANT: ROUGEULLE, CLAIRE
; APPLICANT: AYNER, PHILIP R.
; TITLE OF INVENTION: IDENTIFICATION OF NEURAL DEFECTS ASSOCIATED WITH THE
; TITLE OF INVENTION: NUCLEOSOMAL ASSEMBLY PROTEIN 112 GENE
; FILE REFERENCE: 03495-0203-00000
; CURRENT APPLICATION NUMBER: US/09/847,665
; CURRENT FILING DATE: 2001-10-12

Db	3200	TAAGT	GAGATCA	ATATTTCTTTGGAAAAA	AAATCTTGGGAAAC	AAACCCACAGGGTTTTCGC	3259
QY	220	ATTCCCT	TATCTTCTCTCGCGAA	GTGGTTAAGTCATCTTTAGAGAA	ATATACAGGATA	279	
Db	3260	TGTTGTG	TTTCTCTTTTCTATTTTGT	TGTTTACTT	TAGTCCCTTAGCTAGT	GGATTTAATT	3319
QY	280	TGGA	CTGTATGTGCA	TTCGGAGTTAAACAGTCCAAAGGAGATCTAGAGCA	TTTGGCAA	339	
Db	3320	TTGTTG	TGCGTCTTCATTTTGCA	ATAACAATGCA	TAGATTTAA	AAACTTTGGATGCTTA	3379
QY	340	AGTTCA	ATTTGGCGACAACAT	AAGTGCTGACAAAATCAT	CTTTGGCTTAATAACAGGCT	399	
Db	3380	AGAGCC	TGCATATAGATAG	AAATTTTCAGGCAAA	AACTACATTTATTTGTTAATA	CAGCTT	3439
QY	400	GTATTT	TGGCTCTCTCTATT	420			
Db	3440	GTTCA	TAGGCTCTTGATTTT	3460			

RESULT 12
US-09-974-300-1446
; Sequence 1446, Application US/09974300
; Patent No. US20020146721A1

```

? GENERAL INFORMATION:
? APPLICANT: Berka, Randy M.
? APPLICANT: Clausen, Ib Groth
? TITLE OF INVENTION: Methods For Monitoring Multiple Gene
? TITLE OF INVENTION: Expression
? FILE REFERENCE: 10085-500-US
? CURRENT APPLICATION NUMBER: US/09/974,300
? CURRENT FILING DATE: 2001-10-05
? PRIOR APPLICATION NUMBER: 09/680,598
? PRIOR FILING DATE: 2000-10-06
? PRIOR APPLICATION NUMBER: 60/279,526
? PRIOR FILING DATE: 2001-03-27
? NUMBER OF SEQ ID NOS: 8481
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 1446
? LENGTH: 567
? TYPE: DNA
? ORGANISM: Bacillus licheniformis
US-09-974-300-1446

```

Query Match 1.0%; Score 37.6; DB 10; Length 567;
Best Local Similarity 46.8%; Pred. No. 2.1;
Matches 118; Conservative 0; Mismatches 134; Indels 0; Gaps

Qy	1341	ATTATCTCTGAAGACATAGATAACTCTTCTTCGGTGGTCTTTTGGTTGATGACGAGTGGGAGA	1400
Db	1	ATGCTATTATTGAAGAATTAGATGTTCCTTTCTTTGTGTCGATGACGTGCAGCTCAAGCT	60
Qy	1401	AACCTGTATTCTACAGACTTATTACCAAAAGCAAGTACTGGAAAGTGGTGCAGGACCAACA	1460
Db	61	ATCTCGCACAAAGAATTAATATCAAAAGGCATACACAAATTTGGTCTCATTTCAAAA	120
Qy	1461	TCATGACGAGGATCTTATCAACTCTCGGAAAGGCTTTGTAATTTGGTGATAAAAAATTTG	1520
Db	121	TGGATGATTTACAAGGAAAGTATCGAATGAAGGGGTATATAAAGCGCTTTGGTGAAGCCA	180
Qy	1521	AATTTCTTGCAATTTTCATCGACGACAGTTGCGGGATAATTCAGTGTGGATGTTTGCATCAA	1580
Db	181	AATTAGATTTTACCCGGAGCATGTGCTTTCGTAATATAGACAGCAAGCTGACCTAT	240
Qy	1581	GACCTGGCCTTA	1592
Db	241	ACACAGACCTGA	252

RESULT 13
US-09-864-761-412/c
; Sequence 412, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

Qy 1581 GACCTGGCCTTA 1592
| | | | |
Db 241 ACACAGACCTGA 252

RESULT 13
US-09-864-761-412/C
; Sequence 412, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Weusheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US 09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annotax Sequence Listing Engine vers. 1.1
SEQ ID NO 412
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007543.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
US-09-864-761-412

Query Match 1.0%; Score 36.4; DB 10; Length 499;
Best Local Similarity 53.5%; Pred. No. 4.1;
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 541 TACAGAGGTTCAATTGGCATTGGATTGGAAGAAATTTTCTTTTCTTTTCTAG 600
DB 241 TGCACAGGATTAGTCTTTATAGCCATCAAAAATGATAGTTTTTTCATTTGTTG 182

QY 601 TCGTTCAGCTGACTATAAAGCTTCAGCTTTCATATAGAAAATATATGCGAGGTGTGCTCCA 660
DB 181 TTTTCTAAATGCTATATACATTACCTTCCAGCAAAACACTTTGTGGAAGGTTTAATCTA 122
QY 661 TCGTCCATATGCTCAAAATGCT 682
DB 121 CTACCCAGAAAGCTAGAGAGCT 100
RESULT 14
US-09-939-980-194
Sequence 194, Application US/09939980
Patent No. US20020082234A1
GENERAL INFORMATION:
APPLICANT: Black, Michael
Burnham, Martin
Hodgson, John
Knowles, David
Lonetto, Michael
Nicholas, Richard
Pratt, Julie
Reichard, Richard
Rosenberg, Martin
Ward, Judith
TITLE OF INVENTION: No. US20020082234A1 Polypeptides and Their Uses
NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/939,980
FILING DATE: 27-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/936,165
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 3191 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-09-939-980-194

Query Match 1.0%; Score 36.4; DB 10; Length 3191;
Best Local Similarity 49.0%; Pred. No. 13;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 260 TTAGAGAAATATACAGGATATGGAAGTGTATGCTATGAGGTTAAACAGTCCAAAGGA 319
DB 2823 TTAGAAARACACATTATTTATGAAGATGTTGGACACCGGAGGATATTGAAAAAATAT 2882
QY 320 GGATCTAGAGCAATTTGCCAAAGTTCATTTGCCGACACATAAGTGTGACAAATCATC 379

Db 2883 CGNCTCATCGTGGCAATATATGGTGTCTGTCAGATATAAAAGAAAACAAAGGATTT 2942
QY 380 ACTTGGCTAATACAGGCTGTATTTGGCTCTTCTTCTTATTTGAAGGCTTGGAAATGAAA 439
Db 2943 ACCTTTCCCTAAAGAAAGTCAGTATTTTGAAGAACTTGACTTTAGGTGGATCAGNAAAT 3002
QY 440 ACTGATATTGTC 451
Db 3003 CCTGGTGGTGGC 3014

RESULT 15

US-09-764-847-1401/c
; Sequence 1401, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1401
; LENGTH: 1196
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1401

Query Match 1.08; Score 36.2; DB 10; Length 1196;
Best Local Similarity 50.9%; Pred.No.8;
Matches 138; Conservative 0; Mismatches 128; Indels 5; Gaps 2;
QY 3109 AGCTTTTGACTACAAAGCTGAATATGACAAAGCTGGGTAAATTTAATGGACTACTATG- 3167
Db 315 ACCAATTCATGAAAATGGAACTGTGAAATTTGCTCTTTAAATCAGTACTCTCTGT 256
QY 3168 --GCATAAAAACAGAGGCTGAAATACTAGTGGGCATTTAAGGCATCAAAAACCTTT 3225
Db 255 TCATTTAAAAATAAAATATATAGTTCCTTAAGTTGTATTTGCTAAAGGCATCAGAAGAA 196
QY 3226 TCACCGCAGAAAAGATGCTGAGGCCATAGTGTGCTGTGAGGCCCTTGAGCAGGAGGC 3285
Db 195 CATCTTAAGCAATAGCAATGTGACTATTTTGCCTTATGAGATGGTGTATTTGGGACTAGG 136
QY 3286 AAGAGCCTGGTTCAGAGGCGCTAATGATAT--AGATGACATGTTACCAAGGCTTCGGCT 3343
Db 135 ATGGCCATATATCAATATGTACAGGATATCGAGATCCAAATGTTAGCCAGAGATCTCAAT 76
QY 3344 TGGTACCAGGTTACATATCATCCTACATATT 3374
Db 75 TAGTCCCAAGTTTGTCTCTCATATTCATTTT 45

Search completed: November 5, 2002, 23:26:24
Job time : 283 secs


```
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
BASE COUNT      650 a  435 c  530 g  584 t
ORIGIN
Query Match      19.3%; Score 720.4; DB 11; Length 2199;
Best Local Similarity 64.1%; Pred. No. 2.3e-173;
Matches 1203; Conservative 0; Mismatches 601; Indels 72; Gaps 5;
QY 1601 GATATAGAGCTTCGATGGGTGATTTTCGAGATCAAGATGTCGCAAAATATGCTGCC 1660
D 1601 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 GACATAGGAAGTGGATGGGGAGCTTCGAGATATCAGAAATGGCAAGATGCTGCA 62
QY 1661 AGACTGGTCAATCTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTAGGCATGAGATT 1720
D 1661 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
63 ATACTTGGCAATCTTCAGTCTCCTCAACAGAAACTTTAAAGTACACAAATCTGAGGTG 122
QY 1721 GAAGTTATCCCGATGAAGGTTCAAGGACCAAGCTATGCTTTCTGATGAAATGGT 1780
D 1721 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 GAACGAATCTCTGATATTA---CAATGGCACAAAGTACATATCTCTGATGGAGTTGGA 179
QY 1781 AAAATATCTGTGATCTTGTCTCATAGAGTTGCCCTCAAAATGTGCCCTTCAA---TATACC 1837
D 1781 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 AAGATCTCAGCTAATTTTGCAGTGGAGTGGCTATGAAGTGCAATTTGAACGCTTTGCT 239
QY 1838 CCATCTGTTTTCCAGATCTGTATGGTGGATATGAAGGTGTTGGGGTGTGATCCGGAT 1997
D 1838 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
240 CTTCTCTGTTTTTCAGTAAGGTATGGCGGTTTACAAAGGTGTTCGCTGTAGATACAGA 299
QY 1998 TCATCAATGAAGTGTCTTTTCAGAAAGAGCATGTGCAAAATATGAATCAGACAAATGAAG 1957
D 1998 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 TCAATATCAAGCTTCTTTTCAGAAAGAGCATGTCAAAGTTCCAGTCAGAAATATCACT 359
QY 1958 TTAGATGCTTGGATGAGCAAAATATCAGCCTTGTATCTTAATCTCAACTGATTTACG 2017
D 1958 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
360 CTTGATGCTTGTGATCAGCAAGTACCAACCATGCTTCTGAAATCGCAGTGTGATTA 419
QY 2018 CTCCTGTCTACACTTGGAGTGAAGATGAAGTCTCGAACAAGAGCAAAAGAGAGCTGTA 2077
D 2018 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
420 CTTCTCAACACTTGGGTTAGCGATAATGTCTTTGAGCTAAGCAGAGAGGAGCCTTA 479
QY 2078 GATCAGCTTGATGCTATCTTCATGATCTTTGAAGGCACAGAGGCTTTGGAAATGATG 2137
D 2078 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
480 AGGCAGTTGAACAGAAATGGTAACTGAACACACAGGCTGCTGTAAGCAGTTGAACCTTATG 539
QY 2138 TCTCTGAGAGAACACTTAATATCTCAAGCAATGCTAAACTGTGTTTATAAGCCTGAT 2197
D 2138 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 CCCATGGGAGAGTAACCAATGTAGTTAAAGAAATGTGTCACTGTGCTACCAAGCCTGAT 599
QY 2198 GCTGAGCCTTCTTCTCAATGATGTTGCAACACTTCGCGCATCCAAAGTGTGCTGATTTG 2257
D 2198 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
600 CATGAGCCATATCTTCCATGCTGTCAAAACTTTTAGAGCATCCAAAGCTTCTAGAGTTG 659
QY 2258 CGGACTAGATCAAGATATTTTCCAAATGGAAGCAATGATGGGATGTTTGGATGAA 2317
D 2258 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
660 AAACAAGAGTCAAGATATTCATACAGGGGGGAGCAATGATGGGTGCTGGATGAA 719
QY 2318 TCCAGAACCTTGGATATGCTGAGTGTGTTTTCAGTTTACTGTTGCTGGATGGAGAG 2377
D 2318 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
720 ACCTGCACACTTAAGTAGCGCCAGGTATTCGTCCAAAGTCTTACAGTGCAAGT----- 774
QY 2378 TTTCTGACGATTTACATCCATTTAAACAGAGATCCCAACAGTAATTTTCAATCTG 2437
D 2378 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
775 -----ACCATCGCAAGGCTGTTGTA 794
QY 2438 AAGGAAATGTGGTTGTGCAAAATCCATGCTTGTGATCTGCTGATATCGGTGTTTA 2497
D 2438 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
795 ACTGGAAGAGTAGTTGTGCGCAAAAAATCCTTGTCTCCACCCCTGGTGACATACGGGTTCTC 854
QY 2498 AAGGCTGTAATATGTCGACCGCTGCACACATGGTAGATTGTTGTTATCCCTCAGAAA 2557
D 855 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
855 CAGGCTGTTGATGTTCTCTCTGACCACTTGTGTTGATGTTGTTCTTTCCACAGAG 914
QY 2558 GGAAGAAAGACCTCATCCGAATGAATGTTCTGGAGTGTATTTGGATGGGATATCTACTTT 2617
D 2558 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
915 GGACCAAGGCGCACCTTAATGAGTGTTCAGGAGTGTATCTTGTATGGGACATATATTT 974
QY 2618 GTTGTGGGATCAAGACATGATCCCGCAAGGCAAGTCCAGCGATGGAATATCTCCA 2677
D 2618 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
975 GTTCTTGGGATCCACATCTTATCCAAAGTCTGTTGGTGGATCCCTATGGACTATATCTCA 1034
QY 2678 GCACCCAGCATACAGTTGGACCATGATGTCACAAATGAGGAAGTTGAAGAGTACTTCACC 2737
D 2678 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1035 GCTTCAGCAGAAACATTAGACCATGATGTCATTTTGGAGATACAGGAGTACTTCACA 1094
QY 2738 AACTATATTGTAATGACAGTTTGGAAATCATAGCAAAATGCCCATGTCGTATTTGAGAC 2797
D 2738 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1095 AACTACATATGTAATGAGAGTCTTGGGATTTATGCGCAATGCGCATGTGCTCTTACAGAT 1154
QY 2798 AGAGAACCTGATATGGCCATGAGTATCCATGATCCATCAAAAACTTGTGAGCTCTTTCAATT 2857
D 2798 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1155 CAGGACGATGATGAAGCTGAGAGTCCACCGTGGTTCAACTGCCAAGCTCTTCTCTATA 1214
QY 2858 GAGTGGACTTTCCAAAGACTGTTCCCGCTGAAATPACCAATCAGTTCAGTTCGCCCTTAA 2917
D 2858 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1215 GCTGTGCTATTTCCAAAGACTGAGTGGCGCTGCTGATTCACATGAGCTACATGTCAAG 1274
QY 2918 GAATACCCAGACTTCATGGATGAAGCGGACAGACAGCTATATCTCAGAAAGAGTATT 2977
D 2918 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1275 GAGTATCTCTGACTTCATGAGAACTCGAAGTCAACCTATGATCAAAAGGTTGTGATC 1334
QY 2978 GGAAGACTTTTCAGAAAGTGAAGCAAGCAAGCACTCAGGCTAGCTCTATCGGACCTTC 3037
D 2978 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1335 GGAAGCTCTATAGGAAATTAAGAAAGCACACACCACAC-----ATAAGCACTTC 1385
QY 3038 ACAAGAGATGTTCAAGAGATCATATGATGTGTGATATGGAAGTTGATGGATTTGAAGAT 3097
D 3038 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1386 ACAGGGAAGTGCAGAGCGGCTTTATGACACCGATTTGATTTGTTGATGCTATGAAGAT 1445
QY 3098 TACATTGACGAAGCTTTTGTACTACAAACTGAATATGACAAAGCTGGGTAAATTTAATG 3157
D 3098 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1446 TACATTACTGAGCTATAGATTCAAGGAAGAGTAGATTTCAGGCTGGGTAAATCTATG 1505
QY 3158 GACTACTATGGCATAAAACAGAGGCTGAAATACTTAGTGGGCATTATGAAGGCATCA 3217
D 3158 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1506 GACCACTATGGCATAAAAGTGAAGCTGAGATTAATGAGTGTATTTAAAGATGGCA 1565
QY 3218 AAAAAGCTTTGACGCGCAAGAAAGATGCTGAGGCAATTAGTGTGCTGAGGCGCTTGAGG 3277
D 3218 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1566 AAGAATTTCCACCAAGAGTAGTGTGATGCAATTAAGATGGCGGTGAGATCTTTGAGG 1625
QY 3278 AAGGAGCAGAGCCTTGGTTCAGAGGCGCTA-----ATCATATAGATGACATG 3325
D 3278 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1626 AAAGAAGCTAGGTGCTGGTTCAATGAGATGACAGAGGAGGATGGCCAAAGATGCCATG 1685
QY 3326 TTACCAAGAGCTTCGCGCTTGGTACCAGGTTACATATCATCTACATATTGGGTTGCTAC 3385
D 3326 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1686 GAGGCCAAGSCCTCTGCTGGTACCATTGTTACTTATCATCAGCAGTACTGGGCGAGCTAC 1745
QY 3386 AATCAGGGTTGAAAGAGCTCATTTTCATAGCTTTCCCTGGTGTGTTTATGACAGCTA 3445
D 3386 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1746 AATGAAGGATGATCGCGCGCATCTATTAGCTTCCCATGGTGGCTATATGACAAGCTT 1805
QY 3446 ATCCAGATTAAGAAGG 3461
D 3446 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1806 GTGGCCATCAAGCAGG 1821
D 1806 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 2
BQ507474

```
LOCUS      BQ507474                773 bp      mRNA      linear      EST 22-JUL-2002
DEFINITION EST614889 Generation of a set of potato cDNA clones for microarray
analyses mixed potato tissues Solanum tuberosum cDNA clone STMG059
5' end, mRNA sequence.
ACCESSION  BQ507474
VERSION    BQ507474.2  GI:21923304
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 773)
AUTHORS   Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
            Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
            Karamycheva,S.A.
            Generation of a set of potato cDNA clones for microarray analyses
            Unpublished (2002)
            On Jun 10, 2002 this sequence version replaced gi:21366343.
            Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato@igr.org
            This clone is available through the Research Genetics, contact the
            Research Genetics for further information 1-800-711-6195 or
            cdna@resgen.com
            Seq primer: T3.
FEATURES   Location/Qualifiers
            1..773
             /organism="Solanum tuberosum"
             /cultivar="Kennebec or Binjite"
             /db_xref="taxon:4113"
             /clone="STMG059"
             /clone_lib="Generation of a set of potato cDNA clones for
             microarray analyses mixed potato tissues"
             /tissue_type="mixed tissues"
             /lab_host="SOLR"
             /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
             XhoI; supplier: Combination of untreated and Phytophthora
             infestans-treated libraries of stolons, leaves, leaflets,
             axillary buds of stem explants, petioles, germinating eyes
             tubers, or roots."
BASE COUNT 205 a 125 c 181 g 262 t
ORIGIN
Query Match      17.7%  Score 662;  DB 14;  Length 773;
Best Local Similarity 94.7%  Pred. No. 1.4e-158;
Matches 729;  Conservative 0;  Mismatches 35;  Indels 6;  Gaps 4;

QY  74  ATTTGGTTGCTTCAGTTTCAGTCACTACTGCTGGTAGTGTATTTTTCGATAAATTC 133
Db  8  ATTTGGTTGCTTCAGTTTCAGTCACTACTGCTGGTAGTGTATTTTTCGATAAATTC 67

QY  134 AGGGGGTATTCAGTTGGTG-TTAGCATTTGAAAGTCGAAGTGCACATTTGGAATTTGGCTA 192
Db  68 AGGAGGTATTCAGTTGGTGATAGCATTTGAGCTGGAAGTGCACATTTGGAATTTGGCTG 127

QY  193 CATGGGAAGACAATTCAGGTTTTCGAGTTCCTCTATCTCTCTCGGAAGTGGTTAA 252
Db  128 CATGGGAAGACAATTCAGGTTTTCGAGTTCCTCTCTCTCGGAAGTGGTTAA 187

QY  253 GTCATCTTAGAATAATACAGATATGGAAGTGTATGTCATGTCAGGTTAAACAGTC 312
Db  188 GTCATCTTAGAATAATACAGATATGGAAGTGTATGTCATGTCAGGTTAAACAGTC 247

QY  313 CAAAGGAGGATCTAGAGCATTTGGCAAGTTCAATTTTGGCCGACACATAAGTGTGCACAA 372
Db  248 CAAAGAGGATCTAGAGCATTTGGCAAGTTCAATTTTGGCCGACACATAAGTGTGCACAA 307

QY  373 AATCACTACTTGGCTAATACAGGCTGTATTTTGGCTCTCTCTTATTTGAAGGCTTGGGA 432
Db  308 AATCACTACTTGGCTAATACAGGCTGTATTTTGGCTCTCTCTTATTTGAAGGCTTGGGA 367
```

```
QY  433 AATGAACATGATATGTCCAACTGCGGGGCATATGTGGATCAGATGGATGGCAATACTTT 492
Db  368 AAT---AATGATCTTTGCCAACTGTGGAGGATATGTGGATCAGATGGATGGCAATACTTT 424
QY  493 GAATTTTCGGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAGTACAGAAGTTTC 552
Db  425 GAATTTTCGGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAGTACAGAAGTTTC 484
QY  553 AATCAATTTGGCATTTGGATTTGAAGAAATTTTTCCTTTTATCTAGTGGTTTCAGCTGA 612
Db  485 AATTAATTTGGCATTTGGATTTGAAGAAATTTTTCCTTTTATCTAGTGGTTTCAGCTGA 544
QY  613 CTATAACTTCAGCTTTTCATATCAAAATATA-TGGCAGCTGTGGCCATCGTCCATATG 671
Db  545 TTATAAATTCAGCTTTTCATATCAAAATATA-TGGCAGCTGTGGCCATCGTCCATATG 604
QY  672 GTCAAAATGCTCAGTTTTCCTCATACAGTATTTTGGTCTCTCTCGGATCTTAAAGAGAC 731
Db  605 GTCAAAATGCTCAGTTTTCCTCATACAGTATTTTGGTCTCTCTCGGATCTTAAAGAGAC 664
QY  732 TTGAAAACCTCTGTTTATAGCTTCTTTAAGAAACCTCTGATGATGATGGTGAGACAA 791
Db  665 TTGAAAACCTGTTTATAGCTTCTTTAAGG-AACTCTGATGATGATGGTGAGAGCAG 723
QY  792 CAGATTTCCTCATCTTGGTAGGCTATCTTCTAGCTTATGTTTTCAG 841
Db  724 CAGATTTCCTCATCTTGGTAGGCTATCTTCTAGCTTATGTTTTCAG 773

RESULT 3
BG594667
LOCUS
DEFINITION EST493345 cSTS Solanum tuberosum cDNA clone csts8119 5' sequence,
mRNA sequence.
ACCESSION  BG594667
VERSION    BG594667.1  GI:13612807
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 725)
AUTHORS   van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
            Generations of ESTs from sprouting potato eyes
            Unpublished (2000)
            Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13F-R.
FEATURES   Location/Qualifiers
            1..725
             /organism="Solanum tuberosum"
             /cultivar="Kennebec"
             /db_xref="taxon:4113"
             /clone="csts8119"
             /clone_lib="cSTS"
             /tissue_type="sprouting eyes from tubers"
             /dev_stage="12-14 weeks post harvest"
             /lab_host="SOLR"
             /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
             XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
             taken from tubers. The tubers were incubated at 26C in the
             dark for 2-3 weeks prior to sprouting. The eyes were
             frozen in liquid nitrogen immediately upon removal from
             tubers."
BASE COUNT 228 a 140 c 170 g 187 t
ORIGIN
Query Match      17.7%  Score 661.8;  DB 12;  Length 725;
Best Local Similarity 96.4%  Pred. No. 1.5e-158;
```

JOURNAL COMMENT	FEATURES	SOURCE	BASE COUNT	ORIGIN	Query M Best Lo	Matches
			Qy	3060		1
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	3480
			Qy	3600	541	6000
			Db	3120	61	3180
			Qy	3240	121	3300
			Db	3360	241	

```

RESULT 5
BE435332
LOCUS
DEFINITION EST406410 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
EST406410 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
ACCESSION BE435332
VERSION BE435332.1 GI:9433175
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 618)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
1..618
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG26C9"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCuadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopen accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 196 a 115 c 149 g 157 t
ORIGIN
Query Match 16.2%; Score 605; DB 10; Length 618;
Best Local Similarity 99.7%; Pred. No. 5.4e-144;
Matches 616; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 2801 GAACCTGATATGCCATGAGTATCCATG-CAAAAACCTGCTGAGCTCTTTTCAATTGC 2859
Db 1 GAACCTGATATGCCATGAGTATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGC 60
QY 2860 AGTGGACTTTCCAAAGACTGCTTCCCGTGAATACCATCTCACTTGGCCCTAAAGA 2919
Db 61 AGTGGACTTTCCAAAGACTGCTTCCCGTGAATACCATCTCACTTGGCCCTAAAGA 120
QY 2920 ATACCCAGACTTCATGATAAGCGGACACAGACCATATATCTCAGAAAGATTATTGG 2979
Db 121 ATACCCAGACTTCATGATAAGCGGACACAGACCATATATCTCAGAAAGATTATTGG 180
QY 2980 AAGCTTTTCAGAAAGTGAAGCAACAGACCTCAGGCTAGCTCTATCGGACCTTCAC 3039
Db 181 AAGCTTTTCAGAAAGTGAAGCAACAGACCTCAGGCTAGCTCTATCGGACCTTCAC 240
QY 3040 AAGAGATGTTGCAAGGAGATCATATGCTGATATGGAAGTTGATGATTTGAAGATTA 3099
Db 241 AAGAGATGTTGCAAGGAGATCATATGCTGATATGGAAGTTGATGATTTGAAGATTA 300
QY 3100 CATTGACGAAGCTTTTGACTACAAAACCTGAATATGACACAAAGCTGGGTAATTAATGA 3159
|||||

```

```

Db 301 CATTGACGAAGCTTTTGACTACAAAACCTGAATATGACAAAGCTGGTAAATTAATGA 360
QY 3160 CTACTATGCGATAAAACAGAGGCTGAATACTTTAGTGTGGCATTAATCAAGGCATCAA 3219
Db 361 CTACTATGCGATAAAACAGAGGCTGAATACTTTAGTGTGGCATTAATCAAGGCATCAA 420
QY 3220 AACTTTTGACCGCAGAAAAGATGCTGAGGCCATTAGTGTGTGTGAGGCGCTTGAGGAA 3279
Db 421 AACTTTTGACCGCAGAAAAGATGCTGAGGCCATTAGTGTGTGTGAGGCGCTTGAGGAA 480
QY 3280 GGAGGCAAGAGCGCTGTTCAAGAGGCGTAATGATATAGATGACATGTTACCAAGGCTTC 3339
Db 481 GGAGGCAAGAGCGCTGTTCAAGAGGCGTAATGATATAGATGACATGTTACCAAGGCTTC 540
QY 3340 GGCTTGGTACCCAGGTTACATATCATCTCCTACATATGGGTTGCTACAAATCAGGGTTGAA 3399
Db 541 GGCTTGGTACCCAGGTTACATATCATCTCCTACATATGGGTTGCTACAAATCAGGGTTGAA 600
QY 3400 AAGAGCTCATTTTCATTAG 3417
Db 601 AAGAGCTCATTTTCATTAG 618

RESULT 6
BG597875
LOCUS
DEFINITION ES2496553 cSTS Solanum tuberosum cDNA clone cSTS19C12 5' sequence,
mRNA sequence.
ACCESSION BG597875
VERSION BG597875.1 GI:13616015
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 682)
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,
Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
Generations of ESTs from sprouting potato eyes
Unpublished (2000)
Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
FEATURES
source
1..682
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTS19C12"
/clone_lib="cSTS"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."
BASE COUNT 185 a 108 c 154 g 235 t
ORIGIN
Query Match 16.2%; Score 603.2; DB 12; Length 682;
Best Local Similarity 94.6%; Pred. No. 1.6e-143;
Matches 647; Conservative 0; Mismatches 33; Indels 4; Gaps 2;
QY 74 ATTTGGTGTGCTTCAGTTTCAGTCACACTACTGCTGGGAGTTTATTTTTCATCAACTTC 133
Db 2 ATTTGGTGTGCTTCAGTTTCAGTTTCAGTCACACTACTGCTGGGAGTTTATTTTTCATCAACTTC 61
|||||

```


REFERENCE 1 (bases 1 to 610)
 AUTHORS van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,
 Otterback,R., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning
 C.M., Fry,W.E., Tanksley,S.D. and Baker,B.
 TITLE Generation of ESTs from potato leaves and petioles
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cathy Ronning
 The Institute for Genomic Research
 For clone request: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cdnaresgen.com.
 Location/Qualifiers
 1..610

FEATURES
 source

/organism="Solanum tuberosum"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="CSTB3D19"
 /clone_lib="potato leaves and petioles"
 /tissue_type="leaflets and petioles"
 /dev_stage="8 weeks old plants"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; tissue was supplied by Dr. Fry (Cornell University).
 Leaflets and petioles were isolated from 8 week old
 greenhouse grown plants. The plants were watered and
 fertilized freely. The tissue was immediately frozen in
 liquid nitrogen."

BASE COUNT 181 a 121 c 146 g 162 t

Query Match 15.1%; Score 565.2; DB 12; Length 610;
 Best Local Similarity 96.6%; Pred. No. 8.7e-134;
 Matches 589; Conservative 0; Mismatches 18; Indels 3; Gaps 1;

QY 2480 GGTGATATTCGTTTAAAGGCTTAATGTCGAGCGCTGCACCACATGCGTAGATTGT 2539
 DB 1 GGTGATATTCGTTTAAAGGCTAGATGTTCCAGCGCTGCACCACATGCGTAGATTGT 60
 QY 2540 GTTGATATTCCTCAGAGGAAAGAGACTATCCGATGATGATGTTCTGGGAGTATTG 2599
 DB 61 GTTGATATTCCTCAGAGGAAAGAGACTATCCGATGATGATGTTCTGGGAGTATTG 120
 QY 2600 GATGGGATATCTACTTTGCTGGGATCAAGACATGATCCGCCCAAGCAAGTCCAG 2659
 DB 121 GATGGGATATCTACTTTGCTGGGATCAAGACATGATCCGCCCAAGCAAGTCCAG 180
 QY 2660 CCGATGGAATATCTCCAGACACCCAGCATACAGTTGGACCATGATGTCCACATTTAGGAA 2719
 DB 181 GCGATGGAATATCTCCAGACACCCAGCATACAGTTGGATCATATGTCCACATTTAGGAA 240
 QY 2720 GTTGAAGAGTACTTCCACCAACTATATTGTGATGACAGTTTGGGATCATAGCAATGCC 2779
 DB 241 GTTGAAGAGTACTTCCACCAACTATATTGTGACGACAGTTTGGGATCATAGCAATGCC 300
 QY 2780 CATGTCGATTTGCGACAGAGAACTGATATGGCCATGAGTGATCCATGCAAAAACCTT 2839
 DB 301 CATGTCGATTTGCGACAGAGAACTGATATGGCCATGAGTGATCCATGCAAAAACCTT 360
 QY 2840 GCTGAGCTCTTTCATTCGACGTGACCTTCCAAAGAGTGGTGTCCCGCTGAATACCA 2899
 DB 361 GCTGAGCTCTTTCATTCGACGTGACCTTCCAAAGAGTGGTGTCCCGCTGAATACCA 420
 QY 2900 TCTCAGTTCCGCCCTAAAGAAATACCAGACTTCATGGGATAGCCG---GACAAGACCAAGC 2956
 DB 421 TCTCAGTTCCGCCCTAAAGAAATACCAGACTTCATGGGATAGCCGCCGACAGACACACC 480
 QY 2957 TATATCTCAGAAGAGTTATTGGAAGCTTTTCAGGAAGTGAAGGACAAAGACCTCAG 3016
 DB 481 TATATCTCAGAAGAGTTATTGGAAGCTTTTCAGGAAGTGAAGGACAAAGACCTCAG 540
 QY 3017 GCTAGCTCTATCGACCTTCCAAAGAGATGTTGCAAGGAGATCATATGATGCTGATATG 3076
 DB 541 GCTAGCTCTATCGACCTTCCAAAGGAGTGTTCAGAGAGATCATATGATGCTGATATG 600

QY 3077 GAAGTTGATG 3086
 DB 601 GAAGTTGATG 610

RESULT 9

BE433646 BE433646 570 bp mRNA linear EST 18-MAY-2001
 LOCUS EST400175 tomato breaker fruit, TIGR Lycopersicon esculentum CDNA
 DEFINITION clone cLBE16J9, mRNA sequence.
 ACCESSION BE433646
 VERSION BE433646.1 GI:9431489
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

REFERENCE 1 (bases 1 to 570)

AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
 Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
 Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
 S.D.

TITLE Generation of ESTs from tomato fruit tissue, breaker stage

JOURNAL Unpublished (2000)

COMMENT Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

Location/Qualifiers

1..570

/organism="Lycopersicon esculentum"

/cultivar="FA498"

/db_xref="taxon:4081"

/clone="cLEGI6J9"

/clone_lib="tomato breaker fruit, TIGR"

/tissue_type="Pericarp"

/dev_stage="breaker"

/lab_host="SOLR"

/note="Vector: pBluescript-SKmuadapt; Site 1: EcoRI;

Site 2: XhoI; Fruit were harvested at the breaker stage

(first sign of lycopene accumulation on the blossom end of

the fruit). Fruit were cut in half and the seeds and

locules were discarded prior to freezing the pericarp."

BASE COUNT 157 a 112 c 134 g 167 t

ORIGIN

Query Match 14.6%; Score 545.4; DB 10; Length 570;

Best Local Similarity 99.5%; Pred. No. 1e-128; 1; Indels 2; Gaps 2;

Matches 568; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 2137 GTCTCCTGGAGAGAACACTAATATTCGAAGGCAATGCTAAACTGTGTTTAAAGCTGA 2196

DB 1 GTCTCCTGGAGTGAACACTAATATTCGAAGGCAATGCTAAACTGTGTTTAAAGCTGA 60

QY 2197 TCGTGAGCCCTTCTTTCAATGATGTTGCAACCTTCGGG-CATCCAAAGTGTCTGATT 2255

DB 61 TCGTGAGCCCTTCTTTCAATGATGTTGCAACCTTCGGGCAATGCTGATT 120

QY 2256 TCGGACTAGATCAAGAAATATTTATCCAAATGGAAGAAACATGATGGGATGTTGGATG 2315

DB 121 TCGGACTAGATCAAGAAATATTTATCCAAATGGAAGAAACATGATGGGATG-TTGGATG 179

QY 2316 AATCCGAACCTTGGAAATATGGTCAGGTGTTGTTTCAAGTTTACTGGTGGACATGGAG 2375

DB 180 AATCCGAACCTTGGAAATATGGTCAGGTGTTGTTTCAAGTTTACTGGTGGACATGGAG 239

QY 2376 AGTTTCTGAGATTACATCCATTTAATACAGCAGATCCCAACAGTAATTTCTTC 2435

```

Db 240 AGTTTCTGACGATTACATCCATTTAATAACACGAGATCCACACAGTAATTTTCATTC 299
QY 2436 TGAAGGGAATGTGGTTGTTGCAAAAATCCATCTTGCAATCCCGTGGTGATATTGCTGTTT 2495
Db 300 TGAAGGGAATGTGGTTGTTGCAAAAATCCATCTTGCAATCCCGTGGTGATATTGCTGTTT 359
QY 2496 TAAAGGCTGTAATGTGCGAGCGTGCACACATGGAAGTAGATTGGTTGTTATCCCTCAGA 2555
Db 360 TAAAGGCTGTAATGTGCGAGCGTGCACACATGGAAGTAGATTGGTTGTTATCCCTCAGA 419
QY 2556 AAGGAAAAGACCTCATCCGAATGAATGTTCTGGAGTGATTTGGATGGGATATCTACT 2615
Db 420 AAGGAAAAGACCTCATCCGAATGAATGTTCTGGAGTAGATTGGATGGGATATCTACT 479
QY 2616 TTGTTTGTGGATCAAGACATGATCCGCGCAAGCAAGTCCAGCCGATGGAATATCCTC 2675
Db 480 TTGTTTGTGGATCAAGACATGATCCGCGCAAGCAAGTCCAGCCGATGGAATATCCTC 539
QY 2676 CAGCACCCAGCATACAGTTGGACCATGATGT 2706
Db 540 CAGCACCCAGCATACAGTTGGACCATGATGT 570

RESULT 10
AW932918
LOCUS EST358761 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION
ACCESSION AW932918
VERSION AW932918.1 GI:8108319
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 537)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
Contact: CUGI
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
1..537
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF50H7"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT 177 a 112 c 118 g 130 t
ORIGIN
Query Match 14.2%; Score 529; DB 10; Length 537;
Best Local Similarity 99.1%; Pred. No. 1.6e-124;
Matches 532; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2635 CATGATCCGCGCAAGGCAAGTCCGCGATGGAATATCTCCAGCACCCAGCATACAGTT 2694

```

```

Db 1 CATGATCCGCGCAAGGCAAGTCCGCGATGGAATATCTCCAGCACCCAGCATACAGTT 60
QY 2695 GGACCATGATGTGCACAAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAATGA 2754
Db 61 GGACCATGATGTGCACAAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAATGA 120
QY 2755 CAGTTTGGGAATCATAGCAAAATGCCCATGCTGCTATTTGGCAGACAGAGAACCCTGATATGCC 2814
Db 121 CAGTTTGGGAATCATAGCAAAATGCCCATGCTGCTATTTGGCAGACAGAGAACCCTGATATGCC 180
QY 2815 CATGAGTGATCCATGCAAAAACCTTCTGCTGAGTCTCTTTTCAATTGCAAGTGAATTTCCAAA 2874
Db 181 CATGAGTGATCCATGCAAAAACCTTCTGCTGAGTCTCTTTTCAATTGCAAGTGAATTTCCAAA 240
QY 2875 GACTGTGTTCCCGTGAAATACCATCTCAGTTGCCCTTAAAGATACCCAGACTTCAT 2934
Db 241 GACTGTGTTCCCGTGAAATACCATCTCAGTTGCCCTTAAAGATACCCAGACTTCAT 300
QY 2935 GGATAAGCGGCAAGACAGCATATATCTCAGAAAGAGTTATTTGAAAGCTTTTCAGGAA 2994
Db 301 GGATAAGCGGCAAGACAGCATATATCTCAGAAAGAGTTATTTGAAAGCTTTTCAGGAA 360
QY 2995 AGTAAAGGACAAAGCACCTCAGGCTAGCTCTATPCGCGACCTTCACAAGAGATGTTGCAAG 3054
Db 361 AGTAAAGGACAAAGCACCTCAGGCTAGCTCTATPCGCGACCTTCACAAGAGATGTTGCAAG 420
QY 3055 GAGATCATATGATGCTGATATGGAAGTTGATGAGTTGAAGATTACATTGACGAGCTTT 3114
Db 421 GAGAATCATATGATGCTGATATGGAAGTTGATGAGTTGAAGATTACATTGACGAGCTTT 480
QY 3115 TGACTACAAACTGAATATGACAAACAGCTGGGTAAATTTAATGGACTACTATGGCAT 3171
Db 481 TGACTACAAACTGAATATGACAAACAGCTGGGTAAATTTAATGGACTACTATGGCAT 537

RESULT 11
AW933692
LOCUS EST359535 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION
ACCESSION AW933692
VERSION AW933692.1 GI:8109093
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 525)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
1..525
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT 177 a 112 c 118 g 130 t
ORIGIN
Query Match 14.2%; Score 529; DB 10; Length 537;
Best Local Similarity 99.1%; Pred. No. 1.6e-124;
Matches 532; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2635 CATGATCCGCGCAAGGCAAGTCCGCGATGGAATATCTCCAGCACCCAGCATACAGTT 2694

```


Xhol; cLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"

```

BASE COUNT 151 a 111 c 121 g 142 t
ORIGIN

Query Match 14.1%; Score 525; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 1.7e-123;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTCATCCTGGTGATATCGTGTGTTAAAGCGCTGAATCTTCGAGCGTGA 2523
Db 1 TCCATGCTTCATCCTGGTGATATCGTGTGTTAAAGCGCTGAATCTTCGAGCGTGA 60
QY 2524 CCACATGCTAGATGTTGTTGTTATCCCTCAGAAAGAAAAGACCTCATCCGAATGA 2583
Db 61 CCACATGCTAGATGTTGTTGTTATCCCTCAGAAAGAAAAGACCTCATCCGAATGA 120
QY 2584 TTCTGGGAGTATTTGGATGGGATATCTACTTTGTTGCTGGATCAAGACATGATCCC 2643
Db 121 TTCTGGGAGTATTTGGATGGGATATCTACTTTGTTGCTGGATCAAGACATGATCCC 180
QY 2644 GCCAAGGCAAGTCCAGCGATGGAATATCCTCCAGCACCCAGCATACAGTTGGACCATGA 2703
Db 181 GCCAAGGCAAGTCCAGCGATGGAATATCCTCCAGCACCCAGCATACAGTTGGACCATGA 240
QY 2704 TGTCACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAATGACAGTTGGG 2753
Db 241 TGTCACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAATGACAGTTGGG 300
QY 2754 AATCATAGCAATGCCATGCGTATTGTCAGACAGAGAACCTCATATGGCCATGAGTGA 2823
Db 301 AATCATAGCAATGCCATGCGTATTGTCAGACAGAGAACCTCATATGGCCATGAGTGA 360
QY 2824 TCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGTCAGTGGACCTTCCAAAGACTGTTGT 2883
Db 361 TCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGTCAGTGGACCTTCCAAAGACTGTTGT 420
QY 2884 TCCGCTGAATACCATCTCAGTTGCGCCCTAAAGATACCCAGCTTCATGGATAAGCC 2943
Db 421 TCCGCTGAATACCATCTCAGTTGCGCCCTAAAGATACCCAGCTTCATGGATAAGCC 480
QY 2944 GGACAAGACCATGCTATATCTCAGAAAGAGTTATTGAAAGCTTTT 2988
Db 481 GGACAAGACCATGCTATATCTCAGAAAGAGTTATTGAAAGCTTTT 525

```

RESULT 12
BQ111608
LOCUS
DEFINITION BQ111608 521 bp mRNA linear EST 22-JUL-2002
5' end, mRNA sequence.

ACCESSION BQ111608
VERSION BQ111608.2 GI:21915035
KEYWORDS EST.

SOURCE potato.
ORGANISM Solanum tuberosum

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 521)
Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karamycheva,S.A.

Generation of a set of potato cDNA clones for microarray analyses
Unpublished (2002)

On Apr 17, 2002 this sequence version replaced gi:20163570.
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato@tigr.org

This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cdna@resgen.com

Seq primer: T3.

FEATURES Location/Qualifiers

Source

1..521

/organism="Solanum tuberosum"

/cultivar="Kennebec or Binjite"

/db_xref="taxon:4113"

/clone="STMCB87"

/clone_lib="mixed potato tissues"

/tissue_type="mixed tissues"

/lab_host="SOL6"

/note="Vector: pBluescript SK(-); Site:1: EcoRI; Site:2:

XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes tubers, or roots."

BASE COUNT 155 a 106 c 121 g 139 t

ORIGIN

Query Match 12.9%; Score 482.6; DB 14; Length 521;
Best Local Similarity 96.7%; Pred. No. 1.3e-112;
Matches 504; Conservative 0; Mismatches 14; Indels 3; Gaps 1;

QY 2480 GGTGATATTCGTGTTTAAAGCGCTGTAATGTTTCGAGCGCTGCACACATGTTAGATTGT 2539

Db 1 GGTGATATTCGTGTTTAAAGCGCTGTTAGATGTTCCAGCGCTGCACACATGTTAGATTGT 60

QY 2540 GTTGTATTCCTCAGAAAGAAAAGACCTCATCCGAATGAATGTTCTGGGAGTATTGG 2599

Db 61 GTTGTATTCCTCAGAAAGAAAAGACCTCATCCGAATGAATGTTCTGGGAGTATTGG 120

QY 2600 GATGGGATATCTACTTTGTTGCTGGGATCAAGACATGATCCGCAAGCAAGTCCAG 2659

Db 121 GATGGGATATCTACTTTGTTGCTGGGATCAAGACATGATCCGCAAGCAAGTCCAG 180

QY 2660 CCGATGGAATATCCTCCAGCACCCAGCATACAGTTGGACCATGATGTCACATTTAGGAA 2719

Db 181 CCGATGGAATATCCTCCAGCACCCAGCATACAGTTGGACCATGATGTCACATTTAGGAA 240

QY 2720 GTTGAAGAGTACTTCCAACTATATTGTGAATGACAGTTTGGGAATCATAGCAAAATGCC 2779

Db 241 GTTGAAGAGTACTTCCAACTATATTGTGAATGACAGTTTGGGAATCATAGCAAAATGCC 300

QY 2780 CATGCTGATTTGCAGACAGAACCTGATATGGCCATGAGTGATCCATGCAAAAACCTT 2839

Db 301 CATGCTGATTTGCAGACAGAACCTGATATGGCCATGAGTGATCCATGCAAAAACCTT 360

QY 2840 GCTGAGCTCTTTCAATTCGAGTGGACTTCCAAAAGACTGGTTCCTCCGCTGAATACCA 2899

Db 361 GCTGAGCTCTTTCAATTCGAGTGGACTTCCAAAAGACTGGTTCCTCCGCTGAATACCA 420

QY 2900 TCTCAGTTGGCCCTTAAAGAAATACCCAGACTTTCATGGATAAGCCG---GACAGACGAC 2956

Db 421 TCTCAGTTGGCCCTTAAAGAAATACCCAGACTTTCATGGATAAGCCG---GACAGACGAC 480

QY 2957 TATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGAAAGT 2997

Db 481 TATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGAAAGT 521

RESULT 13

AJ487433

LOCUS

DEFINITION

P3f8, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AJ487433 550 bp mRNA linear EST 30-JUL-2002
AJ487433 Solanum tuberosum cv. Provita Solanum tuberosum cDNA clone
P3f8, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;


```

mRNA sequence.
ACCESSION BG592782.1 GI:13610922
VERSION BG592782.1 GI:13610922
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; eusteroids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 529)
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chieningo,A.,
Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
Generations of ESTs from sprouting potato eyes
Unpublished (2000)
Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source            1..529
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="CSTS2K4"
                     /clone_lib="CSTS"
                     /tissue_types="sprouting eyes from tubers"
                     /dev_stage="12-14 weeks post harvest"
                     /lab_host="SOLR"
                     /note="vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
                     XhoI; various sizes of sprouting eyes (2mm to 15mm) were
                     taken from tubers. The tubers were incubated at 26C in the
                     dark for 2-3 weeks prior to sprouting. The eyes were
                     frozen in liquid nitrogen immediately upon removal from
                     tubers."
BASE COUNT   168 a 103 c 120 g 138 t
ORIGIN
Query Match      12.3%; Score 458.8; DB 12; Length 529;
Best Local Similarity 95.1%; Pred. No. 1.6e-106;
Matches 485; Conservative 0; Mismatches 22; Indels 3; Gaps 1;

QY 2710 AATTGAGGAAGTTGAAGTAGTACTTACCACCACTATTCTGTAATGACAGCTTTGGGAATCAT 2769
DB 20  ACTGAGGAAGTTGAAGTAGTACTTACCACCACTATTCTGTAATGACAGCTTTGGGAATCAT 79

QY 2770 AGCAATGCCCATGTCGATTTCGACAGAGAGACCTGATATGGCCATGATGATCCATG 2829
DB 80  ATCAATGCCCATGTCGATTTCGACAGAGAGACCTGATATGGCCATGATGATCCATG 139

QY 2830 CAAAAAAGCTTGTGAGCTCTTTTCAATTCGAGTGGACCTTCCAAAGACTGTTGTTCCCGC 2889
DB 140 CAAAAAAGCTTGTGAGCTCTGTTCAATTGCGAGTGGACCTTTCGAAAGACTGGTGTCTCTGC 199

QY 2890 TGAATATACCATCTCAGTTTCGCGCCTTAAAGAATACCCAGACTTCATGGATATAGCCG --GA 2946
DB 200 TGAATATACCATCTCAGTTTCGCGCCTTAAAGAATACCCAGACTTCATGGATATAGCCGCGAGA 259

QY 2947 CAAGACGAGCTATATCTCAGAAGAGTATTTCGAAAGCTTTTCAGGAAGTTCAGGAAGTCAAGCAAA 3006
DB 260 CAAGACGAGCTATATCTCAGAAGAGTATTTCGAAAGCTTTTCAGGAAGTTCAGGAAGTCAAGCAAA 319

QY 3007 AGCAGCTCAGGCTAGCTCTCTATCGCGACCTTCACAAAGAGATGTTGCAAGGAGATCATATGA 3066
DB 320 AGCAGCTCAGGCTAGCTCTCTATCGCGACCTTCACAAAGGATGTTGCAAGGAGATCATATGA 379

QY 3067 TGCTGATATGGAAGTTGATGGATTTCAGAGATTACATTCAGGAGCTTTTGACTACAAAC 3126
DB 380 TGCTGATATGGAAGTTGATGGATTTCAGAGATTACATTCAGGAGCTTTTGACTACAAAC 439

QY 3127 TGAATATGACACGAGCTGGGTAATTTTATGAGCTACTATGCGATATGCAATAAAACAGAGGCTGA 3186
DB 440 TGAATATGACACGAGCTGGGTAATTTTATGAGCTACTATGCGATATGCAATAAAACAGAGGCTGA 499

```

```

QY 3187 ATACTTAGTGGTGCATTATGAAGGCATC 3216
|||||
Db 500 AATACTTAGTGGTGCATCATGAAGACATC 529

RESULT 15
AI896465 421 bp mRNA linear EST 18-MAY-2001
LOCUS EST265896 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION cLEC15G23, mRNA sequence.
ACCESSION AI896465
VERSION AI896465
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 421)
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.B.,
Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S., Ronning,
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
    source
    location/Qualifiers
    1..421
    /organism="Lycopersicon esculentum"
    /cultivar="TA496"
    /db_xref="taxon:4081"
    /clone="cLEC15G23"
    /clone.lib="tomato callus, TAMU"
    /tissue_type="callus"
    /dev_stage="25-40 days old"
    /lab_host="XLI-Blue MRF"
    /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato callus EST Library."
BASE COUNT 117 a 85 c 101 g 118 t
ORIGIN
    Query Match 11.3%; Score 421; DB 9; Length 421;
    Best Local Similarity 100.0%; Pred. No. 7.4e-97;
    Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCATGCTTGCATCCTGGTGATTCGTTTAAAGGCTGTAATGTCGAGCGCTGCA 2523
|||||
Db 1 TCATGCTTGCATCCTGGTGATTCGTTTAAAGGCTGTAATGTCGAGCGCTGCA 60

QY 2524 CCACATGGTAGATTGGTTGATTCCTCCAGAAAGGAAAGACCTCATCCGAATGAATG 2583
|||||
Db 61 CCACATGGTAGATTGGTTGATTCCTCCAGAAAGGAAAGACCTCATCCGAATGAATG 120

QY 2584 TTCTGGAGTAGATTGGATGGGATATCTACTTTTGTCTGGGATCAAGACATGATCCC 2643
|||||
Db 121 TTCTGGAGTAGATTGGATGGGATATCTACTTTTGTCTGGGATCAAGACATGATCCC 180

QY 2644 GCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCAGCATACAGTTGGACCATGA 2703
|||||
Db 181 GCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCAGCATACAGTTGGACCATGA 240

QY 2704 TGTCACAATTGAGGAAGTGAAGAGTACTTCCACCAACTATATTGTGAATGCAGTTTGGG 2763
|||||

```

```

Db 241 TGTCACAATTGAGGAAGTGAAGAGTACTTCAACCAACTATATTGTGAATGACAGTTTGGG 300
QY 2764 AATCATAGCAAAATGCCCATGTTCGTATTTCGACAGACAGAGAAACCTGATATGCCCATGAGTGA 2823
|||||
Db 301 AATCATAGCAAAATGCCCATGTTCGTATTTCGACAGACAGAGAAACCTGATATGCCCATGAGTGA 360
QY 2824 TCCATGCAAAAAAAGTTGCTGAGCTCTTTTCAATTGCAAGTGGACTTTCCAAAGACTGGTGT 2883
|||||
Db 361 TCCATGCAAAAAAAGTTGCTGAGCTCTTTTCAATTGCAAGTGGACTTTCCAAAGACTGGTGT 420
QY 2884 T 2884
Db 421 T 421

```

Search completed: November 5, 2002, 23:19:25
Job time : 4565 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2002, 23:52:48 ; Search time 148 Seconds
(without alignments)
1002.981 Million cell updates/sec

Title: US-09-782-874-2
Perfect score: 1114
Sequence: 1 MGKTIQVGFYLLSAFVVK.....RPVNLSSLRQALSHRLVLK 1114

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1114	100.0	1114	22 AAU10006	Tomato RNA-directe
2	1114	100.0	1114	22 AAE00897	Tomato C-protein h
3	218	19.6	218	22 AAE00898	Tomato RNA-directe
4	22	2.0	1125	21 AAB28533	Soybean RNA-direct
5	19	1.7	1108	21 AAB28530	Maize RNA-directe
6	17	1.5	17	22 AAU10009	Tomato RNA-directe
7	17	1.5	17	22 AAE00901	Tomato C-protein-s
8	15	1.3	230	21 AAB28534	Maize RNA-directe
9	14	1.3	14	22 AAU10007	Tomato RNA-directe
10	14	1.3	14	22 AAU10008	Tomato RNA-directe

11 1.3 14 22 AAE00899
12 1.3 14 22 AAE00900
13 1.3 417 21 AAB28529
14 1.1 1196 22 AAG67235
15 1.0 11 22 AAG00903
16 1.0 9 13 22 AAU10010
17 1.0 9 13 22 AAE00902
18 1.0 9 836 21 AAB28532
19 1.0 9 1164 21 AAB28531
20 9 8 60 21 AAB28537
21 8 7 93 22 AAC03533
22 8 7 130 21 AAY70941
23 8 7 283 22 AAU18235
24 8 7 327 21 AAY70947
25 8 7 341 20 AAY09307
26 8 7 341 20 AAY09308
27 8 7 477 22 AAU18155
28 8 7 641 23 AAO21854
29 8 7 830 23 ABB93071
30 8 7 1600 22 ABG09417
31 8 7 1600 22 ABG09723
32 8 7 1600 22 ABG12275
33 8 7 1600 22 ABG20004
34 8 7 2168 22 ABG09420
35 8 7 2168 22 ABG09727
36 8 7 2176 22 ABG12287
37 8 7 2176 22 ABG20022
38 7 0.6 18 22 AAU10011
39 7 0.6 18 22 ABB43113
40 7 0.6 18 22 AAM64015
41 7 0.6 18 22 AAM76838
42 7 0.6 18 22 AAM36944
43 7 0.6 18 22 AAE00904
44 7 0.6 24 22 ABB44400
45 7 0.6 24 22 AAM65478
46 7 0.6 24 22 AAM22018
47 7 0.6 26 20 AAY18363
48 7 0.6 31 20 AAY18362
49 7 0.6 36 21 AAB33978
50 7 0.6 40 21 AAB25008

ALIGNMENTS

RESULT 1
AAU10006
ID AAU10006 standard; Protein; 1114 AA.
XX
AC AAU10006;
XX
DT 08-MAY-2002 (first entry)
XX
DE Tomato RNA-directed RNA polymerase (RdRP) protein.
XX
KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection;
KW immunoprecipitation; immunolocalisation; EC 2.7.7.48; gene therapy.
XX
OS Lycopersicon esculentum.
XX
PN US2001023067-A1.
XX
PD 20-SEP-2001.
XX
PF 08-FEB-2001; 2001US-0782874.
XX
PR 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASENEGGER M.
XX (RIED/) RIEDEL L.

Wassenecker M, Riedel L, Schiebel W, Sanger HL;
WPI; 2001-595798/67.
N-PSDB; AAS17837.

New nucleic acid molecule encoding a polypeptide having the enzymatic activity of RNA-directed RNA polymerase, for modulating gene expression and treating cancer and virus infection in human and animals -

Claim 1; Page 19-22; 34pp; English.

This sequence represents the tomato RNA-directed RNA polymerase (RdRP) protein of the invention. The invention comprises the nucleic acid and protein sequences of RdRP. The protein of the invention can catalyze in vitro transcription of short single stranded RNAs into DNA molecules, or be unprimed. The protein may have cytosolic or virucide activities. The sequences of the invention may be used in gene therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template nucleic acid molecule derived from a nucleic acid molecule which causes a disease are useful for treating a disease caused by the undesired expression or overexpression of a nucleic acid molecule in a human, rat or mouse, by administering the molecules. This system can be used in the preparation of a pharmaceutical composition and for inhibiting expression of any desired gene by transferring the RdRP system to organisms that either lack a comparable mechanism or do not sufficiently express their own RdRP. An antibody or an antagonist or inhibitor to the protein are useful for inhibiting RNA directed RNA synthesis and for ensuring stable heterologous, gene expression in transgenic organisms. The sequence is useful for probes and/or for the control of gene expression, as primers for amplification of nucleic acid molecules and as tools for the detection of expression of the cDNA molecules. Additionally, nucleotide and protein sequences are useful for suppression of undesired gene expression in humans and animals. The RdRP is useful as a therapeutic agent for the control of cancer and virus infection in humans and animals and the antibody is useful for immunoprecipitation or immunolocalisation of the protein, identification of polypeptides interacting with it and screening expression libraries.

Sequence 1114 AA;

Query Match 100.0%; Score 1114; DB 22; Length 1114;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVGFPPYLLSAEYVKSFLKTYGTGTCVCALEVKQSGSRFAKYQFADNISADK 60
DB 1 MGKTIQVGFPPYLLSAEYVKSFLKTYGTGTCVCALEVKQSGSRFAKYQFADNISADK 60

QY 61 IITLANNRLYEGSSYLKAWEMKTDIVQLRAYVDMDGITLNFQCIISDDKFAVLGSTEVS 120
DB 61 IITLANNRLYEGSSYLKAWEMKTDIVQLRAYVDMDGITLNFQCIISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKKFFFLSSGSADYKQLQSYENIWQVVLHRYPGQNAQFLLIQFGAPRIYKRL 180
DB 121 IQFGIGLKKFFFLSSGSADYKQLQSYENIWQVVLHRYPGQNAQFLLIQFGAPRIYKRL 180

QY 181 ENSCYSFKFTPDQWRTTDFPPSWIGLSSSLCLOFRGRVLPNPFEEFFHYAERENNI 240
DB 181 ENSCYSFKFTPDQWRTTDFPPSWIGLSSSLCLOFRGRVLPNPFEEFFHYAERENNI 240

QY 241 TLQTGFTFFYSQKSAIPNVQPPGEGISIPYKILFKISSLVHGCIPGAPALNVYFFFLVDP 300
DB 241 TLQTGFTFFYSQKSAIPNVQPPGEGISIPYKILFKISSLVHGCIPGAPALNVYFFFLVDP 300

QY 301 RRRNVACIEHALKLYIKKCCYDPVWRLTEQYDGYLKGQPKPSITLDDGLVYVRRV 360
DB 301 RRRNVACIEHALKLYIKKCCYDPVWRLTEQYDGYLKGQPKPSITLDDGLVYVRRV 360

QY 361 LVTPCKYVFCGPEVNVSNRVLNRYSEDIDNLRVSVFDEWEKLYSTDLLPKASTGSGVR 420
DB 361 LVTPCKYVFCGPEVNVSNRVLNRYSEDIDNLRVSVFDEWEKLYSTDLLPKASTGSGVR 420

QY 421 TNYIERILTLRKGFVIGDKKEFFLAFSSQLRDNSSVMFASRPGLTANDIRANWGDFSQ 480
DB 421 TNYIERILTLRKGFVIGDKKEFFLAFSSQLRDNSSVMFASRPGLTANDIRANWGDFSQ 480

QY 481 IKNVAKYAARLGSGSSRETLSVLRHETEVIPDVKVHGTSVFSDGIGKISGDFAHVA 540
DB 481 IKNVAKYAARLGSGSSRETLSVLRHETEVIPDVKVHGTSVFSDGIGKISGDFAHVA 540

QY 541 SKCGQYTPSAFIQIRYGGYGVVGPDPSSMKLSLRKSMKSYESDNKLDVLGWSKYQPC 600
DB 541 SKCGQYTPSAFIQIRYGGYGVVGPDPSSMKLSLRKSMKSYESDNKLDVLGWSKYQPC 600

QY 601 YLNRLQITLTLGKGVKDEVLQKQKAEVNDOLDAIILHDSLKAQEALELMSPGENTILKAM 660
DB 601 YLNRLQITLTLGKGVKDEVLQKQKAEVNDOLDAIILHDSLKAQEALELMSPGENTILKAM 660

QY 661 LNCGYKPAEPEFLSMMLQTFRASKLLDLTRGRIIPNGRTMMGCLDSERTLEYGVQVFO 720
DB 661 LNCGYKPAEPEFLSMMLQTFRASKLLDLTRGRIIPNGRTMMGCLDSERTLEYGVQVFO 720

QY 721 FTGAGHGESDDLHPNNSRSTNSNFILKGNVYVAKNCPCLHPGDIRVLRKAVNRALHMY 780
DB 721 FTGAGHGESDDLHPNNSRSTNSNFILKGNVYVAKNCPCLHPGDIRVLRKAVNRALHMY 780

QY 781 DCVVFQKGRPHNECSGSLDGIYFYCWQDDMIPPROVQPMYPPAPSIQLDHDVTI 840
DB 781 DCVVFQKGRPHNECSGSLDGIYFYCWQDDMIPPROVQPMYPPAPSIQLDHDVTI 840

QY 841 EEVEEYFTIYVNDLSGIITANAHVFADEPDMAVSDCKKLAEFLSTAVDPFKTGVPAE 900
DB 841 EEVEEYFTIYVNDLSGIITANAHVFADEPDMAVSDCKKLAEFLSTAVDPFKTGVPAE 900

QY 901 IPSQLRPKXPYDFMDPKDTSYISERVIGKLPKVKDKAPQASSIATTRDVARSYDAD 960
DB 901 IPSQLRPKXPYDFMDPKDTSYISERVIGKLPKVKDKAPQASSIATTRDVARSYDAD 960

QY 961 MEVDGFEDYIDAFYKTEYDNKLGNDYIGIKTEAETLGGIMKASKTFDRRKDAEAI 1020
DB 961 MEVDGFEDYIDAFYKTEYDNKLGNDYIGIKTEAETLGGIMKASKTFDRRKDAEAI 1020

QY 1021 SVAVRALRKEARAFKRRNDIDMLPKASAWHYVHYHPTWGCYNOGLKRAHFISFPWCV 1080
DB 1021 SVAVRALRKEARAFKRRNDIDMLPKASAWHYVHYHPTWGCYNOGLKRAHFISFPWCV 1080

QY 1081 YDQLIQIKDKARNRPNVNLSSLRALQSHRLVVK 1114
DB 1081 YDQLIQIKDKARNRPNVNLSSLRALQSHRLVVK 1114

RESULT 2
AAE00897
ID AAE00897 standard; Protein; 1114 AA.
XX
AC AAE00897;
DT 04-JUL-2001 (first entry)
XX
DE Tomato C-protein having RNA-directed RNA polymerase (RdRP) activity.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein.
XX
OS Lycopersicon esculentum.
XX
FH Key Location/Qualifiers
FT Region 331..346
FT /label= P432
FT /note= "C-protein-specific antibody for
FT immunisation of rabbits"
FT Region 377..390
FT /label= P430
FT /note= "C-protein-specific antibody for
FT immunisation of rabbits"

FT Region 784..797
 FT /label= P433
 FT /note= "C-protein-specific antibody for
 FT immunisation of rabbits"
 FT 1007..1020
 FT /label= P431
 FT /note= "C-protein-specific antibody for
 FT immunisation of rabbits"
 FT US6218142-B1.
 PN 17-APR-2001.
 XX
 XX
 XX 05-MAR-1997; 97US-0811583.
 PF
 XX 05-MAR-1997; 97US-0811583.
 PR
 PA (WASS/) WASENEGGER M.
 PA (RIED/) RIEDEL L.
 XX
 PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
 XX
 XX WPI; 2001-289830/30.
 DR N-PSDB: AAD04370.
 DR
 XX
 XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
 PT polymerase enzymatic activity, useful in modulating gene expression in
 PT plants, humans and animals, as well as in plant cell/tissue cultures or
 PT plant breeding -
 XX
 PS Claim 1; Column 33-40; 31pp; English.
 XX
 CC The present sequence is a tomato C-protein having RNA-directed RNA
 CC polymerase (RdRP) activity. This protein is capable of RNA-directed RNA
 CC synthesis, thus using RNA as a template for synthesizing complementary
 CC RNA molecules. RdRP nucleic acid is useful for modulating gene expression
 CC in plants, humans and animals. This may lead to various physiological,
 CC developmental and/or morphological changes. Transgenic plants containing
 CC RdRP nucleic acid is especially useful in plant cell or tissue cultures
 CC and in plant breeding. RdRP is useful in gene therapy, particularly for
 CC treating a disease that is caused by the undesirable expression or
 CC overexpression of a gene.
 XX
 SQ Sequence 1114 AA;
 Query Match 100.0%; Score 1114; DB 22; Length 1114;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGKTIQVGFYLLSAEVVKSFLKYTGTVCALEVKQSGGSRFAKVOFADNISADK 60
 DB 1 MGKTIQVGFYLLSAEVVKSFLKYTGTVCALEVKQSGGSRFAKVOFADNISADK 60
 QY 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGTTLNFQCISDDKFAVLGSTEVS 120
 DB 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGTTLNFQCISDDKFAVLGSTEVS 120
 QY 121 IQFGIGLKKFFFLSSGADYKQLSYENIQVWLHRLPYGQNAQFLLIQLFGAPRIYKRL 180
 DB 121 IQFGIGLKKFFFLSSGADYKQLSYENIQVWLHRLPYGQNAQFLLIQLFGAPRIYKRL 180
 QY 181 ENSCYSFKEPDPDQWVTTDFPPSWIGLSSSLCQLQFRGVRLENFESFFHYAERENNI 240
 DB 181 ENSCYSFKEPDPDQWVTTDFPPSWIGLSSSLCQLQFRGVRLENFESFFHYAERENNI 240
 QY 241 TIQTGTFEFVSKSALVENVOPPEGISIPYKILFKISSLVQHGCIPGALNVYFFRLVDP 300
 DB 241 TIQTGTFEFVSKSALVENVOPPEGISIPYKILFKISSLVQHGCIPGALNVYFFRLVDP 300
 QY 301 RRRNVACIEHALEKLYIYKECCYDPVRWLTEQYDGYLKGROPPKSPSITLDDGLVYVRRV 360
 DB 301 RRRNVACIEHALEKLYIYKECCYDPVRWLTEQYDGYLKGROPPKSPSITLDDGLVYVRRV 360

QY 361 LVTPCKVYFCGPEVNSRNLRYNSIEDINFLRVSYFVDEEWKLYSTDLLPKASTGSGVR 420
 DB 361 LVTPCKVYFCGPEVNSRNLRYNSIEDINFLRVSYFVDEEWKLYSTDLLPKASTGSGVR 420
 QY 421 TNIIYERILSLRKGFIYGDKKFFFLAFSSQLRDNSVWMEFASRPGLTANDIRAWMGDFSQ 480
 DB 421 TNIIYERILSLRKGFIYGDKKFFFLAFSSQLRDNSVWMEFASRPGLTANDIRAWMGDFSQ 480
 QY 481 IKNVAKYAARLGQSGFSSRRETLISVLRHEIEVIPDVKVHGTSYVFSGIGISGDFAHRA 540
 DB 481 IKNVAKYAARLGQSGFSSRRETLISVLRHEIEVIPDVKVHGTSYVFSGIGISGDFAHRA 540
 QY 541 SKCGLQYTPSAFQIRYGGYKGVGVDPDSSMKLSLRKSKSKYSDNKLKDLVLSWKYQPC 600
 DB 541 SKCGLQYTPSAFQIRYGGYKGVGVDPDSSMKLSLRKSKSKYSDNKLKDLVLSWKYQPC 600
 QY 601 YLNRQLITLLSTLGVKDEVLEQKQKAVDQDAILHDSLKAQALELMSGENTNLIKAM 660
 DB 601 YLNRQLITLLSTLGVKDEVLEQKQKAVDQDAILHDSLKAQALELMSGENTNLIKAM 660
 QY 661 LNCGYKPDABEPTLSMMLQTPRASKLLDLRTRSRIFIPNGRTMMGCLDESTLEYGQVFFVQ 720
 DB 661 LNCGYKPDABEPTLSMMLQTPRASKLLDLRTRSRIFIPNGRTMMGCLDESTLEYGQVFFVQ 720
 QY 721 FTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLAENVRALHHMV 780
 DB 721 FTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLAENVRALHHMV 780
 QY 781 DCVVFQKGRPHNECSGSDLDGDIYFVCWQDMIPPROVQPMYPPAPSIQLDHDTV 840
 DB 781 DCVVFQKGRPHNECSGSDLDGDIYFVCWQDMIPPROVQPMYPPAPSIQLDHDTV 840
 QY 841 EEVEEFTNIVNDSLGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDREKTVPAE 900
 DB 841 EEVEEFTNIVNDSLGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDREKTVPAE 900
 QY 901 IPSQLRKEYPDFMDKPKDTSYISERVIGLFRKVKDKAPQASSIATFTRDVARRSYDAD 960
 DB 901 IPSQLRKEYPDFMDKPKDTSYISERVIGLFRKVKDKAPQASSIATFTRDVARRSYDAD 960
 QY 961 MEVDGPFEDYIDEAFDYKTEYDNKLGNDYXGKTEAEILSGGIMKASKTFDRRKDAEAI 1020
 DB 961 MEVDGPFEDYIDEAFDYKTEYDNKLGNDYXGKTEAEILSGGIMKASKTFDRRKDAEAI 1020
 QY 1021 SVAVRALRKEARAWFKRRNDIDDMPLKASAWYHVTYHTYWGYNQGLKRAHFISFPWCY 1080
 DB 1021 SVAVRALRKEARAWFKRRNDIDDMPLKASAWYHVTYHTYWGYNQGLKRAHFISFPWCY 1080
 QY 1081 YDOLIQIKKDKARNRPVNLSSLRAQLSHRLVLK 1114
 DB 1081 YDOLIQIKKDKARNRPVNLSSLRAQLSHRLVLK 1114
 RESULT 3
 AAEE00898
 ID AAEE00898 standard; Protein; 218 AA.
 XX
 XX AAEE00898;
 XX AC
 XX DT
 XX 04-JUL-2001 (first entry)
 XX
 DE Tomato RNA-directed RNA polymerase 800 (RdRP800) protein.
 XX
 KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein.
 OS
 XX Lycopersicon esculentum.
 XX
 PN US6218142-B1.
 XX
 PD 17-APR-2001.
 XX
 XX 05-MAR-1997; 97US-0811583.
 PF

```

XX 05-MAR-1997; 97US-0811583.
PR
XX (WASS/) WASENEGGER M.
PA (RIED/) RIEDEL L.
XX
XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX WPI; 2001-289830/30.
XX
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding -
XX
XX Example 3; Column 39-42; 31pp; English.
PS
XX The present sequence is tomato RNA-directed RNA polymerase 800
CC (RdRP000) protein which corresponds to position 700-917 of C-protein.
CC C-protein is capable of RNA-directed RNA synthesis, thus using RNA as a
CC template for synthesizing complementary RNA molecules. RdRP nucleic acid
CC is useful for modulating gene expression in plants, humans and animals.
CC This may lead to various physiological, developmental and/or
CC morphological changes. Transgenic plants containing RdRP nucleic acid is
CC especially useful in plant cell or tissue cultures and in plant breeding.
CC RdRP is useful in gene therapy, particularly for treating a disease that
CC is caused by the undesirable expression or overexpression of a gene.
XX
XX Sequence 218 AA;
SQ
Query Match 19.6%; Score 218; DB 22; Length 218;
Best Local Similarity 100.0%; Pred. No. 4.3e-224;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 700 RTMGCLDESRTLEYGVVQFTGAGHGFSDDLHPFNNSRSTNSFILKGNVYVAKNPC 759
DB 1 RTMGCLDESRTLEYGVVQFTGAGHGFSDDLHPFNNSRSTNSFILKGNVYVAKNPC 60
QY 760 LHPGDIRVLKAVNRALHHMVDCVFFQKGRPHNCSGSDLDGDIYFVCWDQDMTPPR 819
DB 61 LHPGDIRVLKAVNRALHHMVDCVFFQKGRPHNCSGSDLDGDIYFVCWDQDMTPPR 120
QY 820 VQPMPEYPPAPSIQLDHDVTIEVEEYFTNYVNDLSLIANAHHVVFADREPDMAMSDPC 879
DB 121 VQPMPEYPPAPSIQLDHDVTIEVEEYFTNYVNDLSLIANAHHVVFADREPDMAMSDPC 180
QY 880 KKLAEFSIAVDFPKTGVPAIPISQLRPKEYPDMFKP 917
DB 181 KKLAEFSIAVDFPKTGVPAIPISQLRPKEYPDMFKP 218
RESULT 4
AAB28533
ID AAB28533 standard; Protein; 1125 AA.
XX
XX AAB28533;
XX
XX 07-FEB-2001 (first entry)
XX
XX Soybean RNA-directed RNA polymerase, SEQ ID NO: 12.
DE
XX Soybean; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
XX Glycine max.
XX
XX WO2000060097-A1.
PN
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09105.
PS
XX 07-APR-1999; 99US-0128094.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase

```

```

XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX WPI; 2000-679376/66.
XX N-PSDB; AAC63742.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance -
XX
XX Claim 10; Page 50-54; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC isolated from plant cDNA libraries. They are useful as probes for
CC genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes.
XX
XX Sequence 1125 AA;
SQ
Query Match 2.0%; Score 22; DB 21; Length 1125;
Best Local Similarity 100.0%; Pred. No. 9.9e-14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 791 RHPNCSGSDLDGDIYFVCWD 812
DB 801 RHPNCSGSDLDGDIYFVCWD 822
RESULT 5
AAB28530
ID AAB28530 standard; Protein; 1108 AA.
XX
XX AAB28530;
XX
XX 07-FEB-2001 (first entry)
XX
XX Maize RNA-directed RNA polymerase, SEQ ID NO: 6.
DE
XX Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
XX Zea mays.
XX
XX WO2000060097-A1.
PN
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09105.
XX
XX 07-APR-1999; 99US-0128094.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX WPI; 2000-679376/66.
XX N-PSDB; AAC63739.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance -
XX
XX Claim 10; Page 36-40; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase

```

CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
 CC isolated from plant cDNA libraries. They are useful as probes for
 CC genetically and physically mapping genes, and as markers for
 CC traits linked to those genes. They are useful for controlling gene
 CC expression and provide mechanisms to engineer plant virus resistance.
 CC They are also useful for plant breeding to develop lines with desired
 CC phenotypes.

XX Sequence 1108 AA;

SQ Query Match 1.7%; Score 19; DB 21; Length 1108;
 Best Local Similarity 100.0%; Pred. No. 1.6e-10; Mismatches 0; Indels 0; Gaps 0;
 Matches 19; Conservative 0;

QY 791 RPHNECSGSDLDGDIYEV 809
 |||||
 Db 794 RPHNECSGSDLDGDIYEV 812
 |||||

RESULT 6

AAU10009
 ID AAU10009 standard; Peptide; 17 AA.

XX AC AAU10009;

XX DT 08-MAY-2002 (first entry)

XX DE Tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P432.

XX KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
 KW transgenic plant; transgenic animal; cancer; viral infection;
 KW immunoprecipitation; immunolocalisation; P432; gene therapy.

XX OS Lycopersicon esculentum.

XX PN US2001023067-A1.

XX PD 20-SEP-2001.

XX PF 08-FEB-2001; 2001US-0782874.

XX PR 05-MAR-1997; 97US-0811583.

XX PA (WASS/) WASSENEGGER M.
 XX PA (RIED/) RIEDEL L.

XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX DR WPI; 2001-595798/67.

XX PT New nucleic acid molecule encoding a polypeptide having the enzymatic
 PT activity of RNA-directed RNA polymerase, for modulating gene expression
 PT and treating cancer and virus infection in human and animals -

XX PS Example 7; Page 14; 34pp; English.

XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
 CC immunogenic peptide P432 used to immunise rabbits and produce an RdRP-
 CC specific antibody. The invention comprises the nucleic acid and
 CC protein sequences of RdRP, the protein of the invention can catalyse in
 CC vitro transcription of short single stranded RNAs into DNA molecules,
 CC this transcription can be either primed by RNA or DNA oligonucleotides
 CC or be unprimed. The protein may have cytosolic or virucide activities.
 CC The sequences of the invention may be used in gene therapy or as an RNA
 CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template
 CC nucleic acid molecule derived from a nucleic acid molecule which causes
 CC a disease are useful for treating a disease caused by the undesired
 CC expression or overexpression of a nucleic acid molecule in a human, rat
 CC or mouse, by administering the molecules. This system can be used in the
 CC preparation of a pharmaceutical composition and for inhibiting
 CC expression of any desired gene by transferring the RdRP system to
 CC organisms that either lack a comparable mechanism or do not sufficiently

CC express their own RdRP. An antibody or an antagonist or inhibitor to the
 CC protein are useful for inhibiting RNA directed RNA synthesis and for
 CC ensuring stable heterologous, gene expression in transgenic organisms.
 CC The sequence is useful for probes and/or for the control of gene
 CC expression, as primers for amplification of nucleic acid molecules and
 CC as tools for the detection of expression of the cDNA molecules.
 CC Additionally, nucleotide and protein sequences are useful for
 CC suppression of undesired gene expression in humans and animals. The RdRP
 CC is useful as a therapeutic agent for the control of cancer and virus
 CC infection in humans and animals and the antibody is useful for
 CC immunoprecipitation or immunolocalisation of the protein, identification
 CC of polypeptides interacting with it and screening expression libraries.

XX SQ Sequence 17 AA;

Query Match 1.5%; Score 17; DB 22; Length 17;
 Best Local Similarity 100.0%; Pred. No. 4.1e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 EYDGYLKGROPPKSPS 347
 |||||
 Db 1 EYDGYLKGROPPKSPS 17
 |||||

RESULT 7

AAE00901
 ID AAE00901 standard; peptide; 17 AA.

XX AC AAE00901;

XX DT 04-JUL-2001 (first entry)

XX DE Tomato C-protein-specific antibody, P432 for immunisation of rabbits.

XX KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
 KW antibody; immunisation.

XX OS Lycopersicon esculentum.

XX PN US6218142-B1.

XX PD 17-APR-2001.

XX PF 05-MAR-1997; 97US-0811583.

XX PR 05-MAR-1997; 97US-0811583.

XX PA (WASS/) WASSENEGGER M.
 XX PA (RIED/) RIEDEL L.

XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX DR WPI; 2001-289830/30.

XX PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
 PT polymerase enzymatic activity, useful in modulating gene expression in
 PT plants, humans and animals, as well as in plant cell/tissue cultures or
 PT plant breeding -

XX PS Example 7; Column 24; 31pp; English.

XX The present sequence is a tomato C-protein-specific antibody, P432 chosen
 CC for immunisation of rabbits. C-protein having RNA-directed RNA polymerase
 CC (RdRP) activity is capable of RNA-directed RNA synthesis, thus using RNA
 CC as a template for synthesising complementary RNA molecules. RdRP nucleic
 CC acid is useful for modulating gene expression in plants, humans and
 CC animals. This may lead to various physiological, developmental and/or
 CC morphological changes. Transgenic plants containing RdRP nucleic acid is
 CC especially useful in plant cell or tissue cultures and in plant breeding.
 CC RdRP is useful in gene therapy, particularly for treating a disease that
 CC is caused by the undesirable expression or overexpression of a gene.

08-MAY-2002 (first entry)

Tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P430.

Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription; cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen; transgenic plant; transgenic animal; cancer; viral infection; immunoprecipitation; immunolocalisation; P430; gene therapy.

Lycopersicon esculentum.

US2001023067-A1.

20-SEP-2001.

08-FEB-2001; 2001US-0782874.

05-MAR-1997; 97US-0811583.

(WASS/) WASSENEGGER M.

(RIED/) RIEDEL L.

Wassenegger M, Riedel L, Schiebel W, Sanger HL;

WPI; 2001-595798/67.

New nucleic acid molecule encoding a polypeptide having the enzymatic activity of RNA-directed RNA polymerase, for modulating gene expression and treating cancer and virus infection in human and animals .

Example 7; Page 14; 34pp; English.

This sequence represents the tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P430 used to immunise rabbits and produce an RdRP-specific antibody. The invention comprises the nucleic acid and protein sequences of RdRP, the protein of the invention can catalyse in vitro transcription of short single stranded RNAs into DNA molecules, this transcription can be either primed by RNA or DNA oligonucleotides or be unprimed. The protein may have cytostatic or virucide activities. The sequences of the invention may be used in gene therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template nucleic acid molecule derived from a nucleic acid molecule which causes a disease are useful for treating a disease caused by the undesired expression or overexpression of a nucleic acid molecule in a human, rat or mouse, by administering the molecules. This system can be used in the preparation of a pharmaceutical composition and for inhibiting expression of any desired gene by transferring the RdRP system to organisms that either lack a comparable mechanism or do not sufficiently express their own RdRP. An antibody or an antagonist or inhibitor to the protein are useful for inhibiting RNA directed RNA synthesis and for ensuring stable heterologous, gene expression in transgenic organisms. The sequence is useful for probes and/or for the control of gene expression, as primers for amplification of nucleic acid molecules and as tools for the detection of expression of the cDNA molecules. Additionally, nucleotide and protein sequences are useful for suppression of undesired gene expression in humans and animals. The RdRP is useful as a therapeutic agent for the control of cancer and virus infection in humans and animals and the antibody is useful for immunoprecipitation or immunolocalisation of the protein. Identification of polypeptides interacting with it and screening expression libraries.

Sequence 14 AA;

Query Match 1.3%; Score 14; DB 22; Length 14;

Best Local Similarity 100.0%; Pred. No. 5.6e-07;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 377 SNRVLRNYSIEDIN 390
|||||

DB 1 SNRVLRNYSIEDIN 14

RESULT 10
AAU10008
ID AAU10008 standard; Peptide; 14 AA.
XX
AC AAU10008;
XX
DT 08-MAY-2002 (first entry)
XX
DE Tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P431.
XX
KW Tonato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection;
KW immunoprecipitation; immunolocalisation; P431; gene therapy.
XX
OS Lycopersicon esculentum.
XX
US2001023067-A1.
XX
PD 20-SEP-2001.
XX
XX 08-FEB-2001; 2001US-0782874.
XX
XX 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASENEGGER M.
XX
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR WPI; 2001-595798/67.
XX
XX New nucleic acid molecule encoding a polypeptide having the enzymatic
PT activity of RNA-directed RNA polymerase, for modulating gene expression
PT and treating cancer and virus infection in human and animals
XX
XX Example 7; Page 14; 34pp; English.
XX
XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
CC immunogenic peptide P431 used to immunise rabbits and produce an RdRP-
CC specific antibody. The invention comprises the nucleic acid and
CC protein sequences of RdRP, the protein of the invention can catalyse in
CC vitro transcription of short single stranded RNAs into DNA molecules,
CC this transcription can be either primed by RNA or DNA oligonucleotides
CC or be unprimed. The protein may have cytosolic or virucide activities.
CC The sequences of the invention may be used in gene therapy or as an RNA
CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template
CC nucleic acid molecule derived from a nucleic acid molecule which causes
CC a disease are useful for treating a disease caused by the undesired
CC expression or overexpression of a nucleic acid molecule in a human, rat
CC or mouse, by administering the molecules. This system can be used in the
CC preparation of a pharmaceutical composition and for inhibiting
CC expression of any desired gene by transferring the RdRP system to
CC organisms that either lack a comparable mechanism or do not sufficiently
CC express their own RdRP. An antibody or an antagonist or inhibitor to the
CC protein are useful for inhibiting RNA directed RNA synthesis and for
CC ensuring stable heterologous, gene expression in transgenic organisms.
CC The sequence is useful for probes and/or for the control of gene
CC expression, as primers for amplification of nucleic acid molecules and
CC as tools for the detection of expression of the cDNA molecules.
CC Additionally, nucleotide and protein sequences are useful for
CC suppression of undesired gene expression in humans and animals. The RdRP
CC is useful as a therapeutic agent for the control of cancer and virus
CC infection in humans and animals and the antibody is useful for
CC immunoprecipitation or immunolocalisation of the protein, identification
CC of polypeptides interacting with it and screening expression libraries.
XX
XX Sequence 14 AA;
SQ
Query Match 1.3%; Score 14; DB 22; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
AAE00899
ID AAE00899 standard; peptide; 14 AA.
XX
AC AAE00899;
XX
DT 04-JUL-2001 (first entry)
XX
DE Tomato C-protein-specific antibody, P430 for immunisation of rabbits.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
KW antibody; immunisation.
XX
OS Lycopersicon esculentum.
XX
US6218142-B1.
XX
PD 17-APR-2001.
XX
XX 05-MAR-1997; 97US-0811583.
XX
XX 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASENEGGER M.
XX
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR WPI; 2001-289830/30.
XX
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding
XX
XX Example 7; Column 24; 31pp; English.
XX
XX The present sequence is a tomato C-protein-specific antibody, P430 chosen
CC for immunisation of rabbits. C-protein having RNA-directed RNA polymerase
CC (RdRP) activity is capable of RNA-directed RNA synthesis, thus using RNA
CC as a template for synthesising complementary RNA molecules. RdRP nucleic
CC acid is useful for modulating gene expression in plants, humans and/or
CC animals. This may lead to various physiological, developmental and/or
CC morphological changes. Transgenic plants containing RdRP nucleic acid is
CC especially useful in plant cell or tissue cultures and in plant breeding.
CC RdRP is useful in gene therapy, particularly for treating a disease that
CC is caused by the undesirable expression or overexpression of a gene.
XX
XX Sequence 14 AA;
SQ
Query Match 1.3%; Score 14; DB 22; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1007 ASKTFDRKDAEAI 1020
DB 1 ASKTFDRKDAEAI 14
XXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXX

RESULT 12
AAE00900
ID AAE00900 standard; peptide; 14 AA.
XX
AC AAE00900;
XX
DT 04-JUL-2001 (first entry)
XX

RESULT 10
AAU10008
ID AAU10008 standard; Peptide; 14 AA.
XX
AC AAU10008;
XX
DT 08-MAY-2002 (first entry)
XX
DE Tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P431.
XX
KW Tonato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection;
KW immunoprecipitation; immunolocalisation; P431; gene therapy.
XX
OS Lycopersicon esculentum.
XX
US2001023067-A1.
XX
PD 20-SEP-2001.
XX
XX 08-FEB-2001; 2001US-0782874.
XX
XX 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASENEGGER M.
XX
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR WPI; 2001-595798/67.
XX
XX New nucleic acid molecule encoding a polypeptide having the enzymatic
PT activity of RNA-directed RNA polymerase, for modulating gene expression
PT and treating cancer and virus infection in human and animals
XX
XX Example 7; Page 14; 34pp; English.
XX
XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
CC immunogenic peptide P431 used to immunise rabbits and produce an RdRP-
CC specific antibody. The invention comprises the nucleic acid and
CC protein sequences of RdRP, the protein of the invention can catalyse in
CC vitro transcription of short single stranded RNAs into DNA molecules,
CC this transcription can be either primed by RNA or DNA oligonucleotides
CC or be unprimed. The protein may have cytosolic or virucide activities.
CC The sequences of the invention may be used in gene therapy or as an RNA
CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template
CC nucleic acid molecule derived from a nucleic acid molecule which causes
CC a disease are useful for treating a disease caused by the undesired
CC expression or overexpression of a nucleic acid molecule in a human, rat
CC or mouse, by administering the molecules. This system can be used in the
CC preparation of a pharmaceutical composition and for inhibiting
CC expression of any desired gene by transferring the RdRP system to
CC organisms that either lack a comparable mechanism or do not sufficiently
CC express their own RdRP. An antibody or an antagonist or inhibitor to the
CC protein are useful for inhibiting RNA directed RNA synthesis and for
CC ensuring stable heterologous, gene expression in transgenic organisms.
CC The sequence is useful for probes and/or for the control of gene
CC expression, as primers for amplification of nucleic acid molecules and
CC as tools for the detection of expression of the cDNA molecules.
CC Additionally, nucleotide and protein sequences are useful for
CC suppression of undesired gene expression in humans and animals. The RdRP
CC is useful as a therapeutic agent for the control of cancer and virus
CC infection in humans and animals and the antibody is useful for
CC immunoprecipitation or immunolocalisation of the protein, identification
CC of polypeptides interacting with it and screening expression libraries.
XX
XX Sequence 14 AA;
SQ
Query Match 1.3%; Score 14; DB 22; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
AAE00899
ID AAE00899 standard; peptide; 14 AA.
XX
AC AAE00899;
XX
DT 04-JUL-2001 (first entry)
XX
DE Tomato C-protein-specific antibody, P430 for immunisation of rabbits.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
KW antibody; immunisation.
XX
OS Lycopersicon esculentum.
XX
US6218142-B1.
XX
PD 17-APR-2001.
XX
XX 05-MAR-1997; 97US-0811583.
XX
XX 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASENEGGER M.
XX
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR WPI; 2001-289830/30.
XX
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding
XX
XX Example 7; Column 24; 31pp; English.
XX
XX The present sequence is a tomato C-protein-specific antibody, P430 chosen
CC for immunisation of rabbits. C-protein having RNA-directed RNA polymerase
CC (RdRP) activity is capable of RNA-directed RNA synthesis, thus using RNA
CC as a template for synthesising complementary RNA molecules. RdRP nucleic
CC acid is useful for modulating gene expression in plants, humans and/or
CC animals. This may lead to various physiological, developmental and/or
CC morphological changes. Transgenic plants containing RdRP nucleic acid is
CC especially useful in plant cell or tissue cultures and in plant breeding.
CC RdRP is useful in gene therapy, particularly for treating a disease that
CC is caused by the undesirable expression or overexpression of a gene.
XX
XX Sequence 14 AA;
SQ
Query Match 1.3%; Score 14; DB 22; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1007 ASKTFDRKDAEAI 1020
DB 1 ASKTFDRKDAEAI 14
XXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXX

RESULT 12
AAE00900
ID AAE00900 standard; peptide; 14 AA.
XX
AC AAE00900;
XX
DT 04-JUL-2001 (first entry)
XX

DE Tomato C-protein-specific antibody, P431 for immunisation of rabbits.
XX
KW Tomato: gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
KW antibody; immunisation.
XX
OS Lycopersicon esculentum.
XX
PN US6218142-B1.
XX
PD 17-APR-2001.
XX
XX 05-MAR-1997; 9705-0811583.
XX
XX 05-MAR-1997; 9705-0811583.
XX
XX (WASS/) WASSENEGGER M.
XX (RIED/) RIEDEL L.
XX
XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX WPI; 2001-289830/30.
XX
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
XX polymerase enzymatic activity, useful in modulating gene expression in
XX plants, humans and animals, as well as in plant cell/tissue cultures or
XX plant breeding.
XX
XX Example 7; Column 24; 31pp; English.
XX
XX The present sequence is a tomato C-protein-specific antibody, P431 chosen
XX for immunisation of rabbits. C-protein having RNA-directed RNA polymerase
XX (RdRP) activity is capable of RNA-directed RNA synthesis, thus using RNA
XX as a template for synthesising complementary RNA molecules. RdRP nucleic
XX acid is useful for modulating gene expression in plants, humans and
XX animals. This may lead to various physiological, developmental and/or
XX morphological changes. Transgenic plants containing RdRP nucleic acid is
XX especially useful in plant cell or tissue cultures and in plant breeding.
XX RdRP is useful in gene therapy, particularly for treating a disease that
XX is caused by the undesirable expression or overexpression of a gene.
XX
XX Sequence 14 AA;
SQ

Query Match 1.3%; Score 14; DB 22; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1007 ASKTFDRRKDAEAI 1020
DB 1 ASKTFDRRKDAEAI 14

RESULT 13
AAB28529
ID AAB28529 standard; Protein; 417 AA.
XX
AC AAB28529;
XX
XX 07-FEB-2001 (first entry)
XX
XX Maize partial RNA-directed RNA polymerase, SEQ ID NO: 4.
XX
XX Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
XX plant virus resistance; plant breeding.
XX
XX Zea mays.
XX
XX WO2000060097-A1.
XX
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09105.
XX
XX

PR 07-APR-1999; 99US-0128094.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX WPI; 2000-679376/66.
XX N-PSDB; AAC63738.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX for controlling gene expression and providing mechanisms to engineer
XX plant virus resistance.
XX
XX Claim 10; Page 34-35; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX isolated from plant cDNA libraries. They are useful as probes for
XX genetically and physically mapping genes, and as markers for
XX traits linked to those genes. They are useful for controlling gene
XX expression and provide mechanisms to engineer plant virus resistance.
XX They are also useful for plant breeding to develop lines with desired
XX phenotypes.
XX
XX Sequence 417 AA;
SQ

Query Match 1.3%; Score 14; DB 21; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 PSAFOIRYGGYKGV 562
DB 62 PSAFOIRYGGYKGV 75

RESULT 14
AAG67235
ID AAG67235 standard; Protein; 1196 AA.
XX
XX AAG67235;
XX
XX 13-NOV-2001 (first entry)
XX
XX Amino acid sequence of a plant SGS2 polypeptide.
XX
XX SGS2; RNA-dependent RNA polymerase; transgene silencing;
XX transgene stability; crop plant; viral resistance.
XX
XX Arabidopsis thaliana.
XX
XX FR2804128-A1.
XX
XX 27-JUL-2001.
XX
XX 26-JAN-2000; 2000FR-0001007.
XX
XX 26-JAN-2000; 2000FR-0001007.
XX
XX (RHOB-) RHOBIO.
XX PA (INRG) INST NAT RECH AGRONOMIQUE.
XX
XX Beclin C, Elmayer T, Mourrain P, Vaucheret H;
XX WPI; 2001-543303/61.
XX N-PSDB; AAH77704.
XX
XX A new plant SGS2 gene involved in encoding an RNA-dependent RNA
XX polymerase and in transgene silencing, increases transgene stability
XX and expression in transgenic plants when it is inactivated.
XX
XX Claim 24; Fig 1; 46pp; French.
XX

CC The present sequence represents a plant SGS2 polypeptide. SGS2 is an
 CC RNA-dependent RNA polymerase and is involved in transgene silencing.
 CC Inactivation of SGS2 is used to increase transgene stability and
 CC expression in plants, particularly crop plants, especially maize, corn,
 CC barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants.
 CC Overexpression of SGS2 can be used to increase resistance to viral
 CC infection in plants.

XX SQ Sequence 1196 AA;

Query Match 1.1%; Score 12; DB 22; Length 1196;
 Best Local Similarity 100.0%; Pred. No. 0.0052;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 KASAWHYVTHP 1058
 DB 1124 KASAWHYVTHP 1135
 |||||

RESULT 15

AAE00903
 ID AAE00903 standard; peptide; 11 AA.

AC AAE00903;

DT 04-JUL-2001 (first entry)

DE Peptide #1 obtained by micro-sequencing RdRP protein.

XX Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein.

XX Lycopersicon esculentum.

PN US6218142-B1.

PD 17-APR-2001.

PF 05-MAR-1997; 97US-0811583.

PR 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.
 PA (RIED/) RIEDEL L.

PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;

DR WPI; 2001-289830/30.
 DR N-PSDB: AAD04381.

XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
 PT polymerase enzymatic activity, useful in modulating gene expression in
 PT plants, humans and animals, as well as in plant cell/tissue cultures or
 PT plant breeding

XX Example 3; Fig 6; 31pp; English.

XX The present sequence is a peptide encoded by an oligonucleotide which is
 CC used to design RNA-directed RNA polymerase (RdRP)-specific PCR primers.
 CC These PCR primers are used to amplify a cDNA encoding tomato C-protein
 CC having RdRP activity. The peptide is obtained by micro-sequencing RdRP
 CC protein. C-protein is capable of RNA-directed RNA synthesis, thus using
 CC RNA as a template for synthesising complementary RNA molecules. RdRP
 CC nucleic acid is useful for modulating gene expression in plants, humans
 CC and animals. This may lead to various physiological, developmental and/or
 CC morphological changes. Transgenic plants containing RdRP nucleic acid is
 CC especially useful in plant cell or tissue cultures and in plant breeding.
 CC RdRP is useful in gene therapy, particularly for treating a disease that
 CC is caused by the undesirable expression or overexpression of a gene.

XX SQ Sequence 11 AA;

Query Match 1.0%; Score 11; DB 22; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.00072;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 EYPDFMDKPKD 919
 DB 1 EYPDFMDKPKD 11
 |||||

RESULT 16

AAU10010
 ID AAU10010 standard; Peptide; 13 AA.

AC AAU10010;

DT 08-MAY-2002 (first entry)

DE Tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P433.

XX Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
 KW cytotatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
 KW transgenic plant; transgenic animal; cancer; viral infection;
 KW immunoprecipitation; immunolocalisation; P433; gene therapy.

XX Lycopersicon esculentum.

PN US2001023067-A1.

PD 20-SEP-2001.

PF 08-FEB-2001; 2001US-0782874.

PR 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.
 PA (RIED/) RIEDEL L.

PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;

DR WPI; 2001-595798/67.

XX New nucleic acid molecule encoding a polypeptide having the enzymatic
 PT activity of RNA-directed RNA polymerase, for modulating gene expression
 PT and treating cancer and virus infection in human and animals

XX Example 7; Page 14; 34pp; English.

XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
 CC immunogenic peptide P433 used to immunise rabbits and produce an RdRP-
 CC specific antibody. The invention comprises the nucleic acid and
 CC protein sequences of RdRP, the protein of the invention can catalyse in
 CC vitro transcription of short single stranded RNAs into DNA molecules,
 CC this transcription can be either primed by RNA or DNA oligonucleotides,
 CC or be unprimed. The protein may have cytotatic or virucide activities.
 CC The sequences of the invention may be used in gene therapy or as an RNA
 CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template
 CC nucleic acid molecule derived from a nucleic acid molecule which causes
 CC a disease are useful for treating a disease caused by the undesired
 CC expression or overexpression of a nucleic acid molecule in a human, rat
 CC or mouse, by administering the molecules. This system can be used in the
 CC preparation of a pharmaceutical composition and for inhibiting
 CC expression of any desired gene by transferring the RdRP system to
 CC organisms that either lack a comparable mechanism or do not sufficiently
 CC express their own RdRP. An antibody or an antagonist or inhibitor to the
 CC protein are useful for inhibiting RNA directed RNA synthesis and for
 CC ensuring stable heterologous, gene expression in transgenic organisms.
 CC The sequence is useful for probes and/or for the control of gene
 CC expression, as primers for amplification of nucleic acid molecules and
 CC as tools for the detection of expression of the cDNA molecules.

XX Additionally, nucleotide and protein sequences are useful for
 CC suppression of undesired gene expression in humans and animals. The RdRP
 CC is useful as a therapeutic agent for the control of cancer and virus
 CC infection in humans and animals and the antibody is useful for
 CC immunoprecipitation or immunolocalisation of the protein, identification

CC of polypeptides interacting with it and screening expression libraries.

```
XX SQ Sequence 13 AA;
Query Match 0.9%; Score 10; DB 22; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 784 VFQKGRPH 793
Db 1 VFQKGRPH 10
|||||

RESULT 17
AAE00902
ID AAE00902 standard; peptide; 13 AA.
XX AC AAE00902;
XX DT 04-JUL-2001 (first entry)
XX DE Tomato C-protein-specific antibody, P433 for immunisation of rabbits.
XX KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
XX KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
XX KW antibody; immunisation.
XX OS Lycopersicon esculentum.
XX PN US6218142-B1.
XX PD 17-APR-2001.
XX PF 05-MAR-1997; 97US-0811583.
XX PR 05-MAR-1997; 97US-0811583.
XX PA (WASS/) WASSENEGGER M.
XX PA (RIED/) RIEDEL L.
XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX DR WPI; 2001-289830/30.
XX PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
XX PT polymerase enzymatic activity, useful in modulating gene expression in
XX PT plants, humans and animals, as well as in plant cell/tissue cultures or
XX PT plant breeding
XX PS Example 7; Column 24; 31pp; English.
XX CC The present sequence is a tomato C-protein-specific antibody, P433 chosen
XX CC for immunisation of rabbits. C-protein having RNA-directed RNA polymerase
XX CC (RdRP) activity is capable of RNA-directed RNA synthesis, thus using RNA
XX CC as a template for synthesising complementary RNA molecules. RdRP nucleic
XX CC acid is useful for modulating gene expression in plants, humans and
XX CC animals. This may lead to various physiological, developmental and/or
XX CC morphological changes. Transgenic plants containing RdRP nucleic acid is
XX CC especially useful in plant cell or tissue cultures and in plant breeding.
XX CC RdRp is useful in gene therapy, particularly for treating a disease that
XX CC is caused by the undesirable expression or overexpression of a gene.
XX SQ Sequence 13 AA;
Query Match 0.9%; Score 10; DB 22; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 784 VFQKGRPH 793
Db 1 VFQKGRPH 10
|||||

RESULT 18
AAB28532
ID AAB28532 standard; Protein; 836 AA.
XX AC AAB28532;
XX DT 07-FEB-2001 (first entry)
XX DE Rice RNA-directed RNA polymerase, SEQ ID NO: 10.
XX KW Rice; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
XX KW plant virus resistance; plant breeding.
XX OS Oryza sativa.
XX PN WO200060097-A1.
XX PD 12-OCT-2000.
XX PF 06-APR-2000; 2000WO-US09105.
XX PR 07-APR-1999; 99US-0128094.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX DR WPI; 2000-679376/66.
XX DR N-PSDB; AAC63741.
XX PT New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX PT for controlling gene expression and providing mechanisms to engineer
XX PT plant virus resistance
XX PS Claim 10; Page 46-49; 62pp; English.
XX CC The present sequence is part of a plant RNA-directed RNA polymerase
XX CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX CC isolated from plant cDNA libraries. They are useful as probes for
XX CC genetically and physically mapping genes, and as markers for
XX CC traits linked to those genes. They are useful for controlling gene
XX CC expression and provide mechanisms to engineer plant virus resistance.
XX CC They are also useful for plant breeding to develop lines with desired
XX CC phenotypes.
XX SQ Sequence 836 AA;
Query Match 0.9%; Score 10; DB 21; Length 836;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 755 AKNPCLHPGD 764
Db 459 AKNPCLHPGD 468
|||||

RESULT 19
AAB28531
ID AAB28531 standard; Protein; 1164 AA.
XX AC AAB28531;
XX DT 07-FEB-2001 (first entry)
XX DE Maize RNA-directed RNA polymerase, SEQ ID NO: 8.
XX KW Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
XX KW plant virus resistance; plant breeding.
XX OS Zea mays.
XX PN WO200060097-A1.
```

```

XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09105.
XX
XX 07-APR-1999; 99US-0128094.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX WPI; 2000-679376/66.
XX DR N-PSDB; AAC63740.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX PT for controlling gene expression and providing mechanisms to engineer
XX PT plant virus resistance
XX
XX Claim 10; Page 41-45; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX CC isolated from plant cDNA libraries. They are useful as probes for
XX CC genetically and physically mapping genes, and as markers for
XX CC traits linked to those genes. They are useful for controlling gene
XX CC expression and provide mechanisms to engineer plant virus resistance.
XX CC They are also useful for plant breeding to develop lines with desired
XX CC phenotypes.
XX
XX Query Match 0.9%; Score 10; DB 21; Length 1164;
XX Best Local Similarity 100.0%; Pred. No. 0.7;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 755 AKNPCLHPGD 764
XX Db |||||||||
XX 787 AKNPCLHPGD 796
XX
XX RESULT 20
XX AAB28537
XX ID AAB28537 standard; Protein; 60 AA.
XX AC
XX AA828537;
XX
XX DT 07-FEB-2001 (first entry)
XX
XX DE Soybean RNA-directed RNA polymerase, SEQ ID NO: 20.
XX
XX KW Soybean; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
XX KW plant virus resistance; plant breeding.
XX OS Glycine max.
XX
XX PN WO200060097-A1.
XX
XX PD 12-OCT-2000.
XX
XX PF 06-APR-2000; 2000WO-US09105.
XX
XX PR 07-APR-1999; 99US-0128094.
XX
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX
XX PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX PI WPI; 2000-679376/66.
XX DR N-PSDB; AAC63746.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX PT

```

```

PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance
XX
XX Claim 23; Page 60; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX CC isolated from plant cDNA libraries. They are useful as probes for
XX CC genetically and physically mapping genes, and as markers for
XX CC traits linked to those genes. They are useful for controlling gene
XX CC expression and provide mechanisms to engineer plant virus resistance.
XX CC They are also useful for plant breeding to develop lines with desired
XX CC phenotypes.
XX
XX SQ Sequence 60 AA;
XX Query Match 0.8%; Score 9; DB 21; Length 60;
XX Best Local Similarity 100.0%; Pred. No. 0.5;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 891 DFPKGTGVP 899
XX Db |||||||||
XX 15 DFPKGTGVP 23
XX
XX RESULT 21
XX AAO03533
XX ID AAO03533 standard; Protein; 93 AA.
XX AC
XX AA03533;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human polypeptide SEQ ID NO 17425.
XX
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX
XX OS Homo sapiens.
XX
XX PN WO200164835-A2.
XX
XX PD 07-SEP-2001.
XX
XX PF 26-FEB-2001; 2001WO-US04927.
XX
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Drmanac RT;
XX
XX DR WPI; 2001-514838/56.
XX DR N-PSDB; AAI83464.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders
XX
XX Claim 20; SEQ ID NO 17425; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or

```

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 93 AA;

QY Query Match 0.7%; Score 8; DB 22; Length 93;
Db Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 LKKFFFFFL 134
Db 51 LKKFFFFFL 58

RESULT 22

AAAY70941
ID: AAAY70941 standard; Protein; 130 AA.

XX AC AAAY70941;

XX DT 09-AUG-2000 (first entry)

XX DE Corn histidine biosynthetic enzyme HisA from clone p0085.csca119r.

XX KW Corn; histidine biosynthetic enzyme; HisA; clone p0085.csca119r;
KW phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase;
KW herbicide; fungicide.

XX OS Zea mays.

XX PN WO200024867-A2.

XX PD 04-MAY-2000.

XX PF 21-OCT-1999; 99WO-US24698.

XX PR 23-OCT-1998; 98US-0105409.

XX PA (DUPO) DU PONT DE NEMOURS & CO E. I.

XX PI Abell LM, Allen SM, Cahoon RE, Tao Y;

XX WPI; 2000-350713/30.

XX DR N-PSDB; AAD00287.

XX PT Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide
PT isomerase, polypeptide (P), useful for discovering compounds which
PT inhibit isomerase activity and act as herbicides and fungicides -

XX PS Claim 2; Page 38-39; 47pp; English.

XX CC The present sequence is a phosphoribosylformimino-5-
CC aminoimidazole carboxamide ribotide isomerase (HisA) which is a
CC histidine biosynthetic enzyme. The sequence is derived from
CC p0085.csca119r clone isolated from a corn shoot culture cDNA library
CC p0085. The present sequence is useful in altering the
CC levels of HisA enzyme in transformed plant cells.

CC It may also be used for discovering compounds that have the ability
CC to inhibit activity of the isomerase, and therefore act as herbicides
CC and fungicides.

XX SQ Sequence 130 AA;

QY Query Match 0.7%; Score 8; DB 21; Length 130;
Db Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 TSYVFSFG 527

Db 99 TSYVFSFG 106

RESULT 23

AAU18235

ID: AAU18235 standard; Protein; 283 AA.

XX AC AAU18235;

XX DT 21-NOV-2001 (first entry)

XX DE Novel human DNA-binding protein #82.

XX KW Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytostatic.

XX OS Homo sapiens.

XX PN WO200155162-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01305.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226868.

XX PR 22-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

XX PR 30-AUG-2000; 2000US-0228924.

XX PR 01-SEP-2000; 2000US-0229287.

XX PR 01-SEP-2000; 2000US-0229343.

XX PR 01-SEP-2000; 2000US-0229344.

XX PR 01-SEP-2000; 2000US-0229345.

XX PR 05-SEP-2000; 2000US-0229509.

XX PR 05-SEP-2000; 2000US-0229513.

XX PR 06-SEP-2000; 2000US-0230437.

XX PR 06-SEP-2000; 2000US-0230438.

XX PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-02311243.
PR 08-SEP-2000; 2000US-02311244.
PR 08-SEP-2000; 2000US-02311413.
PR 08-SEP-2000; 2000US-02311414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0246517.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI: 2001-465557/50.
XX N-PSDB; AAS29111.
XX
XX Nucleic acid molecules encoding human secreted chromosomal binding
XX proteins, used in preventing, treating or ameliorating a disorder, e.g.
XX Alzheimer's and Parkinson's diseases and cancers -
XX
XX Claim 11; SEQ ID No 220; 561pp; English.
XX
XX The present invention relates to the isolation of novel DNA-binding
XX proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding
XX for these proteins. DNA-binding proteins such as histones, chromo
XX (chromatin organisation modifier) domain proteins, and Y-box binding
XX proteins may contribute to diseases resulting from aberrant DNA
XX organisation and/or gene transcription. The sequences of the invention
XX are useful in screening assays to identify antagonists and/or agonists
XX that may enhance or block activities mediated by DNA-binding proteins.
XX Blockers of DNA-binding proteins may be useful in treating disorders.
XX such as malignant diseases (e.g. cancer), autoimmune disorders
XX (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
XX arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
XX diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
XX disease). The polynucleotide sequences of the invention may also be
XX used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding
XX proteins.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 283 AA;
XX
XX Query Match 0.7%; Score 8; DB 22; Length 283;
XX Best Local Similarity 100.0%; Pred. No. 25;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 739 SRSTNSNF 746
XX |
XX Db 114 SRSTNSNF 121
XX
XX
XX RESULT 24
XX AAY70947
XX ID AAY70947 standard; Protein; 327 AA.
XX
XX AC AAY70947;
XX
XX DT 09-AUG-2000 (first entry)
XX
XX DE Wheat histidine biosynthetic enzyme HisA.
XX

KW Wheat; histidine biosynthetic enzyme; HisA; clone wleln.pk0018.b10;
KW phosphoribosylformimino-5-aminimidazole carboxamide ribotide isomerase;
KW herbicide; fungicide.

XX Triticum aestivum.

OS WC200024867-A2.

XX PD 04-MAY-2000.

XX PF 21-OCT-1999; 99WO-US24698.

XX PR 23-OCT-1998; 98US-0105409.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PI Abell LM, Allen SM, Cahoon RE, Tao Y;

XX DR WPI; 2000-350713/30.

DR N-PSDB; AAD00293.

XX PT Phosphoribosylformimino-5-aminimidazole carboxamide ribotide
PT isomerase, polypeptide (P), useful for discovering compounds which
PT inhibit isomerase activity and act as herbicides and fungicides -

XX PS Claim 2; Page 46-47; 47pp; English.

XX CC The present sequence is a phosphoribosylformimino-5-

CC aminimidazole carboxamide ribotide isomerase (HisA) which is a

CC histidine biosynthetic enzyme. The sequence is derived from

CC wleln.pk0018.b10 clone isolated from a wheat leaf cDNA library wleln.

CC The present sequence is useful in altering the

CC levels of HisA enzyme in transformed plant cells.

CC It may also be used for discovering compounds that have the ability

CC to inhibit activity of the isomerase, and therefore act as herbicides

CC and fungicides.

XX SQ Sequence 327 AA;

Query Match 0.7%; Score 8; DB 21; Length 327;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 TSIVFSDG 527

DB 172 TSIVFSDG 179

RESULT 25

AAV09307

ID AAV09307 standard; Protein; 341 AA.

XX AC AAV09307;

XX DT 09-JUL-1999 (first entry)

XX DE Human Vpr interacting protein.

XX KW Human; Vpr interacting protein; VIP; hVIP; cellular receptor; HIV-1;
KW G2/M phase transition; cell cycle; cancer; hyperproliferative disease;
KW autoimmune disease.

XX OS Homo sapiens.

XX PN WO9919359-A1.

XX PD 22-APR-1999.

XX PF 09-OCT-1998; 98WO-US21432.

XX PR 10-OCT-1997; 97US-0949202.

XX PA (UYPE-) UNIV PENNSYLVANIA.

XX PI Ayyavoo V, Mahalingam S, Patel M, Weiner DB;
XX WPI; 1999-277596/23.
DR N-PSDB; AAX55871.

XX PT Human Vpr interacting protein for, e.g. treatment of cancer
XX Claim 1; Page 51-52; 58pp; English.

XX CC The present sequence represents human Vpr interacting protein (hVIP).
CC DNA encoding hVIP can be used to transform host cells to express the
CC protein. Sequences complementary to the DNA can be used in methods for
CC inhibiting expression of hVIP, treating an individual who has cancer,
CC and for inhibiting hVIP activity in a cell. The hVIP can also be used
CC to treat hyperproliferative diseases and some autoimmune diseases.

XX SQ Sequence 341 AA;

Query Match 0.7%; Score 8; DB 20; Length 341;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137

DB 26 FFFFLSSG 33

RESULT 26

AAV09308

ID AAV09308 standard; Protein; 341 AA.

XX AC AAV09308;

XX DT 09-JUL-1999 (first entry)

XX DE Alternative human Vpr interacting protein.

XX KW Human; Vpr interacting protein; VIP; hVIP; cellular receptor; HIV-1;
KW G2/M phase transition; cell cycle; cancer; hyperproliferative disease;
KW autoimmune disease.

XX OS Homo sapiens.

XX PN WO9919359-A1.

XX PD 22-APR-1999.

XX PF 09-OCT-1998; 98WO-US21432.

XX PR 10-OCT-1997; 97US-0949202.

XX PA (UYPE-) UNIV PENNSYLVANIA.

XX PI Ayyavoo V, Mahalingam S, Patel M, Weiner DB;

XX WPI; 1999-277596/23.

DR N-PSDB; AAX55872.

XX PT Human Vpr interacting protein for, e.g. treatment of cancer
XX Example 1; Page 54-56; 58pp; English.

XX CC The present sequence represents human Vpr interacting protein (hVIP).
CC DNA encoding hVIP can be used to transform host cells to express the
CC protein. Sequences complementary to the DNA can be used in methods for
CC inhibiting expression of hVIP, treating an individual who has cancer,
CC and for inhibiting hVIP activity in a cell. The hVIP can also be used
CC to treat hyperproliferative diseases and some autoimmune diseases.

XX SQ Sequence 341 AA;

Query Match 0.7%; Score 8; DB 20; Length 341;

Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137
 |||||
Db 26 FFFFLSSG 33

RESULT 27
AAU18155
ID AAU18155 standard; Protein; 477 AA.
XX AC AAU18155;
XX DT 21-NOV-2001 (first entry)
XX DE
XX DE Novel human DNA-binding protein #2.
XX Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytoskeletal.
XX OS Homo sapiens.
XX WO200155162-A1.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01305.
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0198874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225211.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226581.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.

PR 28-AUG-2001; 2001WO-EP09892.
XX (FARB) BAYER AG.
XX Tietjen K, Weidler M;
PI WPI; 2002-269010/31.
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX Claim 5; SEQ ID NO 2282; 261pp + Sequence Listing; English.
XX The invention relates to identifying target proteins
CC (AB90790-AB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX SQ Sequence 830 AA;
Query Match 0.7%; Score 8; DB 23; Length 830;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 606 LITLLSTL 613
Db 162 LITLLSTL 169
RESULT 30
ABG09417
ID ABG09417 standard; Protein; 1600 AA.
XX AC
XX ABG09417;
DT 13-FEB-2002 (first entry)
XX DE
XX Novel human diagnostic protein #9408.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS
XX Homo sapiens.
XX PN
XX WO200175067-A2.
XX PD
XX 11-OCT-2001.
XX PF
XX 30-MAR-2001; 2001WO-US08631.
XX PR
XX 31-MAR-2000; 2000US-0540217.
XX PR
XX 23-AUG-2000; 2000US-0649167.
XX PA
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR
XX N-PSDB; AAS73604.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID NO 39776; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1600 AA;
Query Match 0.7%; Score 8; DB 22; Length 1600;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 68 RLYFGSSY 75
Db 434 RLYFGSSY 441
RESULT 31
ABG09723
ID ABG09723 standard; Protein; 1600 AA.
XX AC
XX ABG09723;
DT 13-FEB-2002 (first entry)
XX DE
XX Novel human diagnostic protein #9714.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS
XX Homo sapiens.
XX PN
XX WO200175067-A2.
XX PD
XX 11-OCT-2001.
XX PF
XX 30-MAR-2001; 2001WO-US08631.
XX PR
XX 31-MAR-2000; 2000US-0540217.
XX PR
XX 23-AUG-2000; 2000US-0649167.
XX PA
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR
XX N-PSDB; AAS73910.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID NO 40082; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1600 AA;

Query Match 0.7%; Score 8; DB 22; Length 1600;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 RLYFGSSY 75
 |||||
 Db 434 RLYFGSSY 441

RESULT 32
 ABG12275
 ID ABG12275 standard; Protein; 1600 AA.

XX AC ABG12275;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #12266.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR N-PSDB; AAS76462.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX PS Claim 20; SEQ ID No 42634; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1600 AA;

Query Match 0.7%; Score 8; DB 22; Length 1600;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 RLYFGSSY 75
 |||||
 Db 434 RLYFGSSY 441

RESULT 33
 ABG20004
 ID ABG20004 standard; Protein; 1600 AA.

XX AC ABG20004;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #1995.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR N-PSDB; AAS84191.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX PS Claim 20; SEQ ID No 50363; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1500 AA;

Query Match 0.7%; Score 8; DB 22; Length 1600;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75
 |||||
 Db 434 RLYFGSSY 441

RESULT 34

ABG09420
 ID ABG09420 standard; Protein; 2168 AA.

XX AC ABG09420;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #9411.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS73607.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX PS Claim 20; SEQ ID NO 39779; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2168 AA;

Query Match 0.7%; Score 8; DB 22; Length 2168;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75
 |||||
 Db 949 RLYFGSSY 956

RESULT 35

ABG09727
 ID ABG09727 standard; Protein; 2168 AA.

XX AC ABG09727;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #9718.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS73914.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX PS Claim 20; SEQ ID NO 40086; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2168 AA;
 Query Match 0.7%; Score 8; DB 22; Length 2168;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 68 RLYFGSSY 75
 Db 949 RLYFGSSY 956
 |||||

RESULT 36
 ABG12287
 ID ABG12287 standard; Protein; 2176 AA.
 XX AC ABG12287;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #12278.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX PN 11-OCT-2001.
 XX PD 30-MAR-2001; 2001WO-US08631.
 XX PF 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX N-PSDB; AAS76474.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity -
 XX Claim 20; SEQ ID NO 42646; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX and gene mapping, and in recombinant production of (II). The
 XX polynucleotides are also used in diagnostics as expressed sequence tags
 XX for identifying expressed genes. (I) is useful in gene therapy techniques
 XX to restore normal activity of (II) or to treat disease states involving
 XX (II). (II) is useful for generating antibodies against it, detecting or
 XX quantitating a polypeptide in tissue, as molecular weight markers and as
 XX a food supplement. (II) and its binding partners are useful in medical
 XX imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2176 AA;
 Query Match 0.7%; Score 8; DB 22; Length 2176;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 68 RLYFGSSY 75
 Db 949 RLYFGSSY 956
 |||||

RESULT 37
 ABG20022
 ID ABG20022 standard; Protein; 2176 AA.
 XX AC ABG20022;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #20013.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX PN 11-OCT-2001.
 XX PD 30-MAR-2001; 2001WO-US08631.
 XX PF 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX N-PSDB; AAS84209.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity -
 XX Claim 20; SEQ ID NO 50381; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX and gene mapping, and in recombinant production of (II). The
 XX polynucleotides are also used in diagnostics as expressed sequence tags
 XX for identifying expressed genes. (I) is useful in gene therapy techniques
 XX to restore normal activity of (II) or to treat disease states involving
 XX (II). (II) is useful for generating antibodies against it, detecting or
 XX quantitating a polypeptide in tissue, as molecular weight markers and as
 XX a food supplement. (II) and its binding partners are useful in medical
 XX imaging of sites expressing (II). (I) and (II) are useful for treating
 XX disorders involving aberrant protein expression or biological activity.
 XX The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 2176 AA;

Query Match 0.7%; Score 8; DB 22; Length 2176;
 Best Local Similarity 100.0%; Pred. No. 1.8e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 RLYFGSSV 75
 |||||
 Db 949 RLYFGSSV 956

RESULT 38
 AAU10011
 ID AAU10011 standard; Peptide; 18 AA.
 XX AC AAU10011;
 XX DT 08-MAY-2002 (first entry)
 XX DE Tomato RNA-directed RNA polymerase (RdRP) 130 kDa-15 AS sequence.
 XX KW Tomato: RdRP: RNA-directed RNA polymerase; in vitro transcription;
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
 KW transgenic plant; transgenic animal; cancer; viral infection;
 KW immunoprecipitation; immunolocalisation; 130 kDa-15; gene therapy.
 XX OS Lycopersicon esculentum.
 XX FH Key Location/Qualifiers
 FT Misc-difference 15
 FT /label= Asn, Lys
 XX PN US2001023067-A1.
 XX PD 20-SEP-2001.
 XX PF 08-FEB-2001; 2001US-0782874.
 XX PR 05-MAR-1997; 97US-0811583.
 XX PA (WASS/) WASSENEGGER M.
 XX PA (RIED/) RIEDEL L.
 XX PI Wassenecker M, Riedel L, Schiebel W, Sanger HL;
 XX WPI; 2001-595798/67.
 XX DR N-PSDB; AAS17849.
 XX PT New nucleic acid molecule encoding a polypeptide having the enzymatic
 PT activity of RNA-directed RNA polymerase, for modulating gene expression
 PT and treating cancer and virus infection in human and animals -
 XX Example 3; Fig 6; 34pp; English.
 XX This sequence represents a tomato RNA-directed RNA polymerase (RdRP)
 CC peptide sequence used to create a degenerate nucleotide sequence and
 CC degenerate PCR primers used to amplify the RdRP cDNA sequence of the
 CC invention. The invention comprises the nucleic acid and protein
 CC sequences of RdRP, the protein of the invention can catalyze in vitro
 CC transcription of short single stranded RNAs into DNA molecules,
 CC this transcription can be either primed by RNA or DNA oligonucleotides
 CC or be unprimed. The protein may have cytosolic or virucide activities.
 CC The sequences of the invention may be used in gene therapy or as an RNA
 CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template

CC nucleic acid molecule derived from a nucleic acid molecule which causes
 CC a disease are useful for treating a disease caused by the undesired
 CC expression or overexpression of a nucleic acid molecule in a human, rat
 CC or mouse, by administering the molecules. This system can be used in the
 CC preparation of a pharmaceutical composition and for inhibiting
 CC expression of any desired gene by transferring the RdRP system to
 CC organisms that either lack a comparable mechanism or do not sufficiently
 CC express their own RdRP. An antibody or an antagonist or inhibitor to the
 CC protein are useful for inhibiting RNA directed RNA synthesis and for
 CC ensuring stable heterologous, gene expression in transgenic organisms.
 CC The sequence is useful for probes and/or for the control of gene
 CC expression, as primers for amplification of nucleic acid molecules and
 CC as tools for the detection of expression of the cDNA molecules.
 CC Additionally, nucleotide and protein sequences are useful for
 CC suppression of undesired gene expression in humans and animals. The RdRP
 CC is useful as a therapeutic agent for the control of cancer and virus
 CC infection in humans and animals and the antibody is useful for
 CC immunoprecipitation or immunolocalisation of the protein, identification
 CC of polypeptides interacting with it and screening expression libraries.
 XX SQ Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 640 KAQEALE 646
 |||||
 Db 1 KAQEALE 7

RESULT 39
 ABB43113
 ID ABB43113 standard; Peptide; 18 AA.
 XX AC ABB43113;
 XX DT 04-FEB-2002 (first entry)
 XX DE Peptide #10619 encoded by human foetal liver single exon probe.
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX OS Homo sapiens.
 XX PN WO200157277-A2.
 XX PD 09-AUG-2001.
 XX PR 30-JAN-2001; 2001WO-US00669.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX Claim 27; SEQ ID NO 35749; 639pp + sequence listing; English.
 XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 572 KLSLRKS 578
Db 12 KLSLRKS 18
|||||||

RESULT 40
AAM64015
ID AAM64015 standard; Protein; 18 AA.
XX
XX AC AAM64015;
AC AAM64015;
XX
XX DT 05-NOV-2001 (first entry)
DT 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 36120.
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 36120.
XX
XX KW Human: brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
XX OS Homo sapiens.
OS Homo sapiens.
XX
XX PN WO200157275-A2.
PN WO200157275-A2.
XX
XX PD 09-AUG-2001.
PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00667.
PF 30-JAN-2001; 2001WO-US00667.
XX
XX PR 04-FEB-2000; 2000US-0180312.
PR 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-0234687.
PR 21-SEP-2000; 2000US-0234687.
XX
XX PR 27-SEP-2000; 2000US-0236359.
PR 27-SEP-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000GB-0024253.
PR 04-OCT-2000; 2000GB-0024253.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-483445/52.
DR WPI; 2001-483445/52.
XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
PT Single exon nucleic acid probes for analyzing gene expression in human
XX
XX PS Example 4; SEQ ID NO: 36120; 650pp + Sequence Listing; English.
PS Example 4; SEQ ID NO: 36120; 650pp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
XX Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 572 KLSLRKS 578
Db 12 KLSLRKS 18
|||||||

RESULT 42
AAM36944
ID AAM36944 standard; Protein; 18 AA.
XX
XX AC AAM36944;
AC AAM36944;
XX
XX DT 17-OCT-2001 (first entry)
DT 17-OCT-2001 (first entry)
XX
XX DE Peptide #10981 encoded by probe for measuring placental gene expression.
DE Peptide #10981 encoded by probe for measuring placental gene expression.
XX

Qy 572 KLSLRKS 578
Db 12 KLSLRKS 18
|||||||

RESULT 41
AAM76838
ID AAM76838 standard; Protein; 18 AA.
XX
XX AC AAM76838;
AC AAM76838;
XX
XX DT 06-NOV-2001 (first entry)
DT 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 37144.
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 37144.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
OS Homo sapiens.
XX
XX PN WO200157276-A2.
PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00668.
PF 30-JAN-2001; 2001WO-US00668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
PR 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-0234687.
PR 21-SEP-2000; 2000US-0234687.
XX
XX PR 27-SEP-2000; 2000US-0236359.
PR 27-SEP-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000GB-0024253.
PR 04-OCT-2000; 2000GB-0024253.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488900/53.
DR WPI; 2001-488900/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX PS Example 4; SEQ ID NO: 37144; 658pp + Sequence Listing; English.
PS Example 4; SEQ ID NO: 37144; 658pp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
XX SQ Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 572 KLSLRKS 578
Db 12 KLSLRKS 18
|||||||

RESULT 42
AAM36944
ID AAM36944 standard; Protein; 18 AA.
XX
XX AC AAM36944;
AC AAM36944;
XX
XX DT 17-OCT-2001 (first entry)
DT 17-OCT-2001 (first entry)
XX
XX DE Peptide #10981 encoded by probe for measuring placental gene expression.
DE Peptide #10981 encoded by probe for measuring placental gene expression.
XX

KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.

XX Homo sapiens.

PN WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta.

XX Claim 27; SEQ ID No 37213; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;
XX see AA131315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.

SQ Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 KLSLRKS 578

DB 12 KLSLRKS 18

RESULT 43

AAE00904

ID AAE00904 standard; peptide; 18 AA.

XX AAE00904;

XX 04-JUL-2001 (first entry)

XX Peptide #2 obtained by micro-sequencing Rdrp protein.

XX Tomato; gene therapy; RNA-directed RNA polymerase; Rdrp; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein.

XX Lycopersicon esculentum.

XX Key Location/Qualifiers

FT Misc-difference 6

FT /note= "Encoded by YTN"

FT Misc-difference 9

FT /note= "Encoded by WSN"

FT Misc-difference 15

FT /label= Lys, Asn

FT /note= "Encoded by AAN"

FT Misc-difference 17

FT /note= "Encoded by YTN"

XX US6218142-B1.

XX 17-APR-2001.

XX 05-MAR-1997; 97US-0811583.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASSENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-289830/30.

XX N-PSDB; AAD04382.

XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
XX polymerase enzymatic activity, useful in modulating gene expression in
XX plants, humans and animals, as well as in plant cell/tissue cultures or
XX plant breeding.

XX Example 3; Fig 6; 31pp; English.

XX The present sequence is a peptide encoded by an oligonucleotide which is
XX used to design RNA-directed RNA polymerase (Rdrp)-specific PCR primers.
XX These PCR primers are used to amplify a cDNA encoding tomato C-protein
XX having Rdrp activity. The peptide is obtained by micro-sequencing Rdrp
XX protein. C-protein is capable of RNA-directed RNA synthesis, thus using
XX RNA as a template for synthesizing complementary RNA molecules. Rdrp
XX nucleic acid is useful for modulating gene expression in plants, humans
XX and animals. This may lead to various physiological, developmental and/or
XX morphological changes. Transgenic plants containing Rdrp nucleic acid is
XX especially useful in plant cell or tissue cultures and in plant breeding.
XX Rdrp is useful in gene therapy, particularly for treating a disease that
XX is caused by the undesirable expression or overexpression of a gene.

SQ Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 640 KAQEALE 646

DB 1 KAQEALE 7

RESULT 44

ABB4400

ID ABB4400 standard; Peptide; 24 AA.

XX ABB4400;

XX 04-FEB-2002 (first entry)

XX Peptide #11906 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WC200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX PS 27-SEP-2000; 2000US-0236359.
XX PA 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 27; SEQ ID NO 37035; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 24 AA;
XX Query Match 0.6%; Score 7; DB 22; Length 24;
XX Best Local Similarity 100.0%; Pred. No. 29; Mismatches 0; Indels 0; Gaps 0;
XX Matches 7; Conservative 0;
XX QY 209 LSSSLCL 215
XX Db 7 LSSSLCL 13
XX
XX RESULT 45
XX AAM65478
XX ID AAM65478 standard; Protein; 24 AA.
XX AC AAM65478;
XX DT 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 37583.
XX Human; brain expressed exon; gene expression analysis; probe;
XX Microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer.
XX Homo sapiens.
XX WO200157275-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00667.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -

```

XX SQ Sequence 24 AA;
Query Match 0.6%; Score 7; DB 22; Length 24;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 209 LSSSLCL 215
Db 7 LSSSLCL 13

RESULT 47
AAY18363
ID AAY18363 standard; peptide; 26 AA.
XX AC AAY18363;
XX DT 20-AUG-1999 (first entry)
XX DE Human chromogranins A fragment.
XX KW Chromogranins A; CgA; human; antibody production; detection.
XX OS Homo sapiens.
XX PN JP11153598-A.
XX PD 08-JUN-1999.
XX PF 20-NOV-1997; 97JP-0336452.
XX PR 20-NOV-1997; 97JP-0336452.
XX PA (YANA-) YANAIHARA KENYUSHO KK.
XX WPI; 1999-390410/33.
New antibody - has reactivity specific to human chromogranins A
(CgA)
Claim 1; Page 2; 10pp; Japanese.
This sequence represents a fragment of human chromogranins A (CgA). The
invention relates to an antibody having a reactivity specific to human
CgA, which is prepared by immunising a warm-blooded animal, other than
human, using this peptide. The antibody is useful in a method for
detecting and/or measuring human CgA. The method can measure and detect
human CgA with a high sensitivity.
XX SQ Sequence 31 AA;
Query Match 0.6%; Score 7; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 568 DSSMKLS 574
Db 8 DSSMKLS 14
|||||
RESULT 49
AAB33978
ID AAB33978 standard; Peptide; 36 AA.
XX AC AAB33978;
XX DT 02-FEB-2001 (first entry)
XX DE Human secreted protein encoded by CDNA #17.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX OS Homo sapiens.
XX PN WO200056765-A1.
XX PD 28-SEP-2000.
XX PF 16-MAR-2000; 2000WO-US06923.
XX PR 19-MAR-1999; 99US-0125364.
XX PR 08-DEC-1999; 99US-0169623.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-502215/57.
XX N-PSDB; AAC59408.
Nucleic acid molecules encoding human secreted proteins, used in
preventing, treating or ameliorating a disorder, e.g. Alzheimer's and

```

PT Parkinson's diseases and cancers -
XX Claim 11; Page 363; 410pp; English.
XX Sequences AAB33963-B34006 represent the amino acid sequences of 48
CC human secreted proteins encoded by the genes AAC59332-59439. The genes
CC and proteins are useful for preventing, ameliorating or treating
CC medical conditions, e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and
CC ulcerative colitis; (c) cardiovascular disorders such as myocardial
CC ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral
CC anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections.
XX Sequence 36 AA;
SQ Query Match 0.6%; Score 7; DB 21; Length 36;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 130 FFFFLSS 136
Db 20 FFFFLSS 26
RESULT 50
RAB25008
ID AAB25008 standard; Peptide; 40 AA.
XX AAB25008;
XX 27-NOV-2000 (first entry)
XX Plant SDF encoded polypeptide sequence SEQ List 2 NO:38.
XX Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;
KW SDF; genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control.
XX Plant.
XX WO2000040695-A2.
XX 13-JUL-2000.
XX 07-JAN-2000; 2000WO-US00466.
XX 08-JAN-1999; 99US-0115293.
XX (CERE-) CERES INC.
XX Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L;
XX WPI; 2000-465970/40.
XX New corn plant and Arabidopsis thaliana sequence-determined DNA
PT fragments, useful for expressing gene products and for controlling
PT expression of a target gene -
XX Claim 14; Page 615; 673pp; English.
XX The present invention describes polynucleotides, such as complete cDNA
CC sequences and/or sequences of genomic DNA encompassing complete genes,
CC portions of genes, and/or intergenic regions, collectively referred to
CC as sequence-determined DNA fragments (SDFs), from corn plants and

CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a
CC target gene, either as a promoter, a structural gene, an UTR or as a
CC 3' termination sequence. They are also useful as tools for genetic
CC mapping, and identification of a particular individual plant or for
CC clustering a group of plants with a common trait. AAB78433 to AAB78630
CC and AAB24605 to AAB25099 represent the specifically claimed
CC polynucleotide sequences and polypeptides encoded by them given in the
XX present invention.
SQ Sequence 40 AA;
Query Match 0.6%; Score 7; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 130 FFFFLSS 136
Db 19 FFFFLSS 25
Search completed: November 6, 2002, 03:33:28
Job time : 164 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:36:48 ; Search time 83 Seconds
(without alignments)

394.905 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 1114

Sequence: 1 MCKTIQVFGFPYLLSAEVVK.....RPVLNLSLRAQLSHRLVLIK 1114

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : Issued Patents_Aa.*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1114	100.0	1114	4	US-08-811-583-2
2	218	19.6	218	4	US-08-811-583-3
3	17	1.5	17	4	US-08-811-583-12
4	14	1.3	14	4	US-08-811-583-10
5	14	1.3	14	4	US-08-811-583-11
6	10	0.9	13	4	US-08-811-583-13
7	8	0.7	341	3	US-08-949-202-2
8	8	0.7	341	3	US-08-949-202-4
9	8	0.7	341	4	US-09-418-175-2
10	8	0.7	341	4	US-09-418-175-4
11	8	0.7	341	4	US-08-529-245-2
12	8	0.7	341	4	US-09-529-245-4
13	8	0.7	506	4	US-09-457-040B-17
14	7	0.6	149	4	US-09-134-001C-3012
15	7	0.6	154	4	US-09-134-001C-3132
16	7	0.6	165	4	US-09-455-960-7
17	7	0.6	220	4	US-08-297-431B-31
18	7	0.6	221	4	US-08-297-431B-2
19	7	0.6	221	4	US-08-297-431B-4
20	7	0.6	221	4	US-08-297-431B-6
21	7	0.6	221	4	US-08-297-431B-8
22	7	0.6	221	4	US-08-297-431B-10
23	7	0.6	221	4	US-08-297-431B-12
24	7	0.6	221	4	US-08-297-431B-14
25	7	0.6	221	4	US-08-297-431B-16
26	7	0.6	221	4	US-08-297-431B-18
27	7	0.6	221	4	US-08-297-431B-20

28	7	0.6	320	1	US-08-565-386-17	Sequence 17, Appl
29	7	0.6	322	2	US-08-576-626A-33	Sequence 33, Appl
30	7	0.6	394	4	US-08-791-115B-27	Sequence 27, Appl
31	7	0.6	403	4	US-08-791-115B-1	Sequence 1, Appl
32	7	0.6	403	4	US-08-791-115B-23	Sequence 23, Appl
33	7	0.6	403	4	US-08-791-115B-25	Sequence 25, Appl
34	7	0.6	430	4	US-08-791-115B-7	Sequence 7, Appl
35	7	0.6	470	2	US-08-933-821-6	Sequence 6, Appl
36	7	0.6	470	3	US-08-960-507-6	Sequence 6, Appl
37	7	0.6	470	4	US-09-136-828-6	Sequence 6, Appl
38	7	0.6	470	4	US-09-332-928A-6	Sequence 6, Appl
39	7	0.6	470	4	US-09-136-801-6	Sequence 6, Appl
40	7	0.6	470	4	US-09-332-929-6	Sequence 6, Appl
41	7	0.6	501	3	US-08-968-563-15	Sequence 15, Appl
42	7	0.6	501	4	US-08-969-683A-15	Sequence 15, Appl
43	7	0.6	501	4	US-09-297-928-11	Sequence 11, Appl
44	7	0.6	565	4	US-08-906-156A-12	Sequence 12, Appl
45	7	0.6	577	4	US-09-183-266A-49	Sequence 49, Appl
46	7	0.6	645	4	US-08-791-115B-6	Sequence 6, Appl
47	7	0.6	742	4	US-08-791-115B-5	Sequence 5, Appl
48	7	0.6	834	4	US-09-143-571-29	Sequence 29, Appl
49	7	0.6	884	4	US-09-741-150-4	Sequence 4, Appl
50	7	0.6	971	2	US-08-724-354D-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-811-583-2
; Sequence 2, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1114 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-583-2

Query Match

100.0%; Score 1114; DB 4; Length 1114;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVFGPPYLLSABVWSFKLEKTYGTVCALVEKSKGSRFAKQVQFADNISADK 60
DB 1 MGKTIQVFGPPYLLSABVWSFKLEKTYGTVCALVEKSKGSRFAKQVQFADNISADK 60

QY 61 IITLANNRLYFGSSYLKAWEMKTDIVOLRAYVQDMGDTILNFCQISDDKFAVLGSTEVS 120
DB 61 IITLANNRLYFGSSYLKAWEMKTDIVOLRAYVQDMGDTILNFCQISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKKFFFLSSGADYKQLSVENIOWVHLRPGVQNAQFLLIOLFGAPRIYKRL 180
DB 121 IQFGIGLKKFFFLSSGADYKQLSVENIOWVHLRPGVQNAQFLLIOLFGAPRIYKRL 180

QY 181 ENSCYSPFKETPDQWVRTTDFPSSWIGLSSSLCLOFRGRVRLNPFEEFFHYAERENNI 240
DB 181 ENSCYSPFKETPDQWVRTTDFPSSWIGLSSSLCLOFRGRVRLNPFEEFFHYAERENNI 240

QY 241 TLOGTFFVYSKALPNVQPPGEGISIPYKILFKISLVOHGCIPGALNVYFFRLVDP 300
DB 241 TLOGTFFVYSKALPNVQPPGEGISIPYKILFKISLVOHGCIPGALNVYFFRLVDP 300

QY 301 RRRNVACIEHALEKLYIKKCCYDPVRLTEQYDGLKGRQPKPSITLDDGLVYVRRV 360
DB 301 RRRNVACIEHALEKLYIKKCCYDPVRLTEQYDGLKGRQPKPSITLDDGLVYVRRV 360

QY 361 LVTPCKYVFCGPEVNVNRLVRYSEDIDNLFVRSFVDEWEKLYSTDLLPKASTGSGYR 420
DB 361 LVTPCKYVFCGPEVNVNRLVRYSEDIDNLFVRSFVDEWEKLYSTDLLPKASTGSGYR 420

QY 421 TNYIERILSLRKGFVIGDKKFFFLAFSSQLRDNVSMFASRPLGTANDIRAWMGDFSQ 480
DB 421 TNYIERILSLRKGFVIGDKKFFFLAFSSQLRDNVSMFASRPLGTANDIRAWMGDFSQ 480

QY 481 IKNVAKYARLQSGFSGSRSTLSVLRHEIEVDPVKVHGTSYVFSGIGKISGDFAHRA 540
DB 481 IKNVAKYARLQSGFSGSRSTLSVLRHEIEVDPVKVHGTSYVFSGIGKISGDFAHRA 540

QY 541 SKCGLQTPSAFQIRYGGYGVGVDPDSSMKLSLRKMSKYESDNIKLDVLGWSKYQPC 600
DB 541 SKCGLQTPSAFQIRYGGYGVGVDPDSSMKLSLRKMSKYESDNIKLDVLGWSKYQPC 600

QY 601 YLNQLITLSTLVQKVEVLEQKQEAVDQDAILHDSLKAQALELMSPGENTNLIKAM 660
DB 601 YLNQLITLSTLVQKVEVLEQKQEAVDQDAILHDSLKAQALELMSPGENTNLIKAM 660

QY 661 LNCGYKPADPFLSMQLQTRASKLLDLRTSRIFIPNGRTMGCCLDESRTLEYGVQVQ 720
DB 661 LNCGYKPADPFLSMQLQTRASKLLDLRTSRIFIPNGRTMGCCLDESRTLEYGVQVQ 720

QY 721 FTGAGHGEFSDDLHPFNNSRSTNSFILKGNVYVAKNPLCPHGDIVRLKAVNVRALHHV 780
DB 721 FTGAGHGEFSDDLHPFNNSRSTNSFILKGNVYVAKNPLCPHGDIVRLKAVNVRALHHV 780

QY 781 DCVVFQPKGRPHNECSGSLDGDIIYFVCWQDMIPROVQPMEXPPAPSIQLDHDTV 840
DB 781 DCVVFQPKGRPHNECSGSLDGDIIYFVCWQDMIPROVQPMEXPPAPSIQLDHDTV 840

QY 841 EEVEEYFTNYIVNDSGLIIANAHVVFADREPDMAMSDPCCKKLAEFLSIADVDFKTVPAE 900
DB 841 EEVEEYFTNYIVNDSGLIIANAHVVFADREPDMAMSDPCCKKLAEFLSIADVDFKTVPAE 900

QY 901 IPSQLRPKEYPDMKDPKTSYISERVIGKLFVKVKQKAPQASSIAFTTRDVARSSDAD 960
DB 901 IPSQLRPKEYPDMKDPKTSYISERVIGKLFVKVKQKAPQASSIAFTTRDVARSSDAD 960

QY 961 MEVDGFEDYIDEADYKTEYDNKLGNDYIGIKTEAELLSSGIMKASKTFDRKDAEAI 1020
DB 961 MEVDGFEDYIDEADYKTEYDNKLGNDYIGIKTEAELLSSGIMKASKTFDRKDAEAI 1020

QY 1021 SVAVRALKEARAWKRRNDIDMLPKASANYHYTYHTYWCYNQGLKRAHFISFPWCV 1080
DB 1021 SVAVRALKEARAWKRRNDIDMLPKASANYHYTYHTYWCYNQGLKRAHFISFPWCV 1080

DB 1021 SVAVRALKEARAWKRRNDIDMLPKASANYHYTYHTYWCYNQGLKRAHFISFPWCV 1080

QY 1081 YDQLIQIKKDKARNRPVLNLSLRAQLSHRLVLK 1114
DB 1081 YDQLIQIKKDKARNRPVLNLSLRAQLSHRLVLK 1114

RESULT 2
US-08-811-583-3
; Sequence 3, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wasseuegger, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sangster, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-811-583-3

Query Match 19.6%; Score 218; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 7.4e-211;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 RTMMGCLDSRTLEYGVQVFTGAGHGEFSDDLHPFNNSRSTNSFILKGNVYVAKNPC 759
DB 1 RTMMGCLDSRTLEYGVQVFTGAGHGEFSDDLHPFNNSRSTNSFILKGNVYVAKNPC 60

QY 760 LHPGDIRVLRKAVNVRALHHVDCVVFQPKGRPHNECSGSLDGDIIYFVCWQDMIPPR 819
DB 61 LHPGDIRVLRKAVNVRALHHVDCVVFQPKGRPHNECSGSLDGDIIYFVCWQDMIPPR 120

QY 820 QVQPMYPPAPSIQLDHDTVIEVEEYFTNYIVNDSGLIIANAHVVFADREPDMAMSDPC 879
DB 121 QVQPMYPPAPSIQLDHDTVIEVEEYFTNYIVNDSGLIIANAHVVFADREPDMAMSDPC 180

QY 880 KKLAEFLSIADVDFKTVPAEIPSQLRPKEYPDMKDP 917
DB 181 KKLAEFLSIADVDFKTVPAEIPSQLRPKEYPDMKDP 218

```
RESULT 3
US-08-811-583-12
; Sequence 12, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RRPR)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-811-583-12
Query Match 1.5%; Score 17; DB 4; Length 17;
Best Local Similarity 100.0%; Pred. No. 8.6e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 331 EQYDGLKGROPPKSPS 347
Db 1 EQYDGLKGROPPKSPS 17

RESULT 4
US-08-811-583-10
; Sequence 10, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RRPR)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
```

```
ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-811-583-10
Query Match 1.3%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 7.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 377 SNEVLNYSIEDIDN 390
Db 1 SNEVLNYSIEDIDN 14

RESULT 5
US-08-811-583-11
; Sequence 11, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RRPR)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
```

; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-811-583-11

Query Match 1.3%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 7.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1007 ASKTFDRKDAEAI 1020
Db 1 ASKTFDRKDAEAI 14

RESULT 6

US-08-811-583-13
; Sequence 13, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMAIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLIMERASE (RGRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/811,583
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-811-583-13

Query Match 0.9%; Score 10; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0073;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 784 VFPQKGRPH 793
Db 1 VFPQKGRPH 10

RESULT 7

US-08-949-202-2
; Sequence 2, Application US/08949202

; Patent No. 6060587
; GENERAL INFORMATION:
; APPLICANT: Weiner, David B
; APPLICANT: Ayyavoo, Velupandi
; APPLICANT: Mahalingam, Sundarasamy
; APPLICANT: Patel, Mamata
; TITLE OF INVENTION: CELLULAR RECEPTOR FOR HIV-1 VPR ESSENTIAL FOR
; TITLE OF INVENTION: G2/M PHASE TRANSITION OF THE CELL CYCLE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6060587rls
; STREET: One Liberty Place 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: Wordperfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/949,202
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPAP-0222
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-949-202-2

Query Match 0.7%; Score 8; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137
Db 26 FFFFLSSG 33

RESULT 8

US-08-949-202-4
; Sequence 4, Application US/08949202
; Patent No. 6060587
; GENERAL INFORMATION:
; APPLICANT: Weiner, David B
; APPLICANT: Ayyavoo, Velupandi
; APPLICANT: Mahalingam, Sundarasamy
; APPLICANT: Patel, Mamata
; TITLE OF INVENTION: CELLULAR RECEPTOR FOR HIV-1 VPR ESSENTIAL FOR
; TITLE OF INVENTION: G2/M PHASE TRANSITION OF THE CELL CYCLE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6060587rls
; STREET: One Liberty Place 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: Wordperfect 6.0

Query Match 0.9%; Score 10; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0073;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7

US-08-949-202-2
; Sequence 2, Application US/08949202

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/949,202
;; FILING DATE:
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Deluca, Mark
;; REGISTRATION NUMBER: 33,229
;; REFERENCE/DOCKET NUMBER: UPAP-0222
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-568-3100
;; TELEFAX: 215-568-3439
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 341 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-949-202-4

Query Match 0.7%; Score 8; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137
| | | | | | | |
Db 26 FFFFLSSG 33

RESULT 9
US-09-418-175-2
; Sequence 2, Application US/09418175
; Patent No. 6172201
; GENERAL INFORMATION:
; APPLICANT: Weiner, David B
; APPLICANT: Ayyavoo, Velpandi
; APPLICANT: Mahalingam, Sundarasamy
; APPLICANT: Patel, Mamata
; TITLE OF INVENTION: CELLULAR RECEPTOR FOR HIV-1 VPR ESSENTIAL FOR
; TITLE OF INVENTION: G2/M PHASE TRANSITION OF THE CELL CYCLE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6172201iris
; STREET: One Liberty Place 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: Wordperfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/418,175
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,202
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPAP-0222
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-418-175-2

Query Match 0.7%; Score 8; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 FFFFLSSG 137
| | | | | | | |
Db 26 FFFFLSSG 33

RESULT 10
US-09-418-175-4
; Sequence 4, Application US/09418175
; Patent No. 6172201
; GENERAL INFORMATION:
; APPLICANT: Weiner, David B
; APPLICANT: Ayyavoo, Velpandi
; APPLICANT: Mahalingam, Sundarasamy
; APPLICANT: Patel, Mamata
; TITLE OF INVENTION: CELLULAR RECEPTOR FOR HIV-1 VPR ESSENTIAL FOR
; TITLE OF INVENTION: G2/M PHASE TRANSITION OF THE CELL CYCLE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6172201iris
; STREET: One Liberty Place 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: Wordperfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/418,175
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,202
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPAP-0222
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-418-175-4

Query Match 0.7%; Score 8; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137
| | | | | | | |
Db 26 FFFFLSSG 33

RESULT 11
US-09-529-245-2
; Sequence 2, Application US/09529245
; Patent No. 6448078
; GENERAL INFORMATION:
; APPLICANT: Weiner, David B
; APPLICANT: Ayyavoo, Velpandi
; APPLICANT: Mahalingam, Sundarasamy

```
; APPLICANT: Patel, Mamata
; TITLE OF INVENTION: Cellular Receptor For HIV-1 VPR Essential for G2/M
; FILE REFERENCE: UPAP0376
; CURRENT APPLICATION NUMBER: US/09/529,245
; CURRENT FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 08/949,202
; PRIOR FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-529-245-2

Query Match      0.7%; Score 8; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 FFFFLSSG 137
Db 26 FFFFLSSG 33

RESULT 12
US-09-529-245-4
; Sequence 4, Application US/09529245
; Patent No. 6448078
; GENERAL INFORMATION:
; APPLICANT: Weiner, David B
; APPLICANT: Ayayoco, Velpandi
; APPLICANT: Mahalingam, Sundarasamy
; APPLICANT: Patel, Mamata
; TITLE OF INVENTION: Cellular Receptor For HIV-1 VPR Essential for G2/M
; FILE REFERENCE: UPAP0376
; CURRENT APPLICATION NUMBER: US/09/529,245
; CURRENT FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 08/949,202
; PRIOR FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-529-245-4

Query Match      0.7%; Score 8; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 FFFFLSSG 137
Db 26 FFFFLSSG 33

RESULT 13
US-09-457-040B-17
; Sequence 17, Application US/09457040B
; Patent No. 6387641
; GENERAL INFORMATION:
; APPLICANT: Vertex Pharmaceuticals Incorporated
; APPLICANT: Bellon, Steve
; TITLE OF INVENTION: Crystallized P38 Complexes
; FILE REFERENCE: VPI/98-14
; CURRENT APPLICATION NUMBER: US/09/457,040B
; CURRENT FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 506
```

```
; TYPE: PRT
; ORGANISM: Yeast
US-09-457-040B-17

Query Match      0.7%; Score 8; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 923 ISERVICK 930
Db 314 ISERVICK 321

RESULT 14
US-09-134-001C-3012
; Sequence 3012, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3012
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3012

Query Match      0.6%; Score 7; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 606 LITLST 612
Db 18 LITLST 24

RESULT 15
US-09-134-001C-3132
; Sequence 3132, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3132
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3132

Query Match      0.6%; Score 7; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 934 KVKDKAP 940
Db 51 KVKDKAP 57
```

RESULT 16
US-09-455-960-7
; Sequence 7, Application US/09455960
; Patent No. 6361776
; GENERAL INFORMATION:
; APPLICANT: Delcayre, Alain
; TITLE OF INVENTION: Compositions isolated from *M. vaccae* and
; TITLE OF INVENTION: their use in modulation of immune responses.
; FILE REFERENCE: 11000.1047
; CURRENT APPLICATION NUMBER: US/09/455.960
; CURRENT FILING DATE: 1999-12-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 165
; TYPE: PRT
; ORGANISM: *Mycobacterium vaccae*
US-09-455-960-7

Query Match 0.6%; Score 7; DB 4; Length 165;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKGGL 545
Db 34 VASKGGL 40

RESULT 17
US-08-297-431B-31
; Sequence 31, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E
; APPLICANT: Gulick, Andrew M
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297.431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 220 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-297-431B-31

Query Match 0.6%; Score 7; DB 4; Length 220;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 124 KDKARNR 130

RESULT 18
US-08-297-431B-2
; Sequence 2, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E
; APPLICANT: Gulick, Andrew M
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297.431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-297-431B-2

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 19
US-08-297-431B-4
; Sequence 4, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E
; APPLICANT: Gulick, Andrew M
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B
; APPLICANT: Kramer, Katharine

```

; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-4

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 20
US-08-297-431B-6
; Sequence 6, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E.
; APPLICANT: Gulick, Andrew M.
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B.
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-8

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 20
US-08-297-431B-6
; Sequence 6, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E.
; APPLICANT: Gulick, Andrew M.
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B.
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-8

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131
```

```

; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-6

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 21
US-08-297-431B-8
; Sequence 8, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E.
; APPLICANT: Gulick, Andrew M.
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B.
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-8

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131
```

Db 125 KDKARNR 131

RESULT 22

US-08-297-431B-10

Sequence 10, Application US/08297431B

Patent No. 6136605

GENERAL INFORMATION:

APPLICANT: Fahl, William E

APPLICANT: Gulick, Andrew M

APPLICANT: Manoharan, T Herbert

APPLICANT: Puchalski, Ralph B

APPLICANT: Kramer, Katharine

APPLICANT: Wasserman Wyeth W

TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saul, Ewing, Remick & Saul, LLP

STREET: 1500 Market Street, 38th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19102-2186

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/297,431B

FILING DATE: August 26, 1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36252

REFERENCE/DOCKET NUMBER: WARF F039

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-972-8386

TELEFAX: 215-972-2292

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 221 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-297-431B-10

Query Match 0.6%; Score 7; DB 4; Length 221;

Best Local Similarity 100.0%; Pred. No. le+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1089 KDKARNR 1095

Db 125 KDKARNR 131

RESULT 23

US-08-297-431B-12

Sequence 12, Application US/08297431B

Patent No. 6136605

GENERAL INFORMATION:

APPLICANT: Fahl, William E

APPLICANT: Gulick, Andrew M

APPLICANT: Manoharan, T Herbert

APPLICANT: Puchalski, Ralph B

APPLICANT: Kramer, Katharine

APPLICANT: Wasserman Wyeth W

TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saul, Ewing, Remick & Saul, LLP

STREET: 1500 Market Street, 38th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19102-2186

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/297,431B

FILING DATE: August 26, 1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36252

REFERENCE/DOCKET NUMBER: WARF F039

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-972-8386

TELEFAX: 215-972-2292

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 221 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-297-431B-12

Query Match 0.6%; Score 7; DB 4; Length 221;

Best Local Similarity 100.0%; Pred. No. le+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1089 KDKARNR 1095

Db 125 KDKARNR 131

RESULT 24

US-08-297-431B-14

Sequence 14, Application US/08297431B

Patent No. 6136605

GENERAL INFORMATION:

APPLICANT: Fahl, William E

APPLICANT: Gulick, Andrew M

APPLICANT: Manoharan, T Herbert

APPLICANT: Puchalski, Ralph B

APPLICANT: Kramer, Katharine

APPLICANT: Wasserman Wyeth W

TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saul, Ewing, Remick & Saul, LLP

STREET: 1500 Market Street, 38th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19102-2186

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/297,431B

FILING DATE: August 26, 1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36252

REFERENCE/DOCKET NUMBER: WARF F039

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-972-8386

TELEFAX: 215-972-2292

INFORMATION FOR SEQ ID NO: 14:

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-297-431B-14
Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 25
US-08-297-431B-16
; Sequence 16, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E
; APPLICANT: Gulick, Andrew M
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-297-431B-16
Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 26
US-08-297-431B-18
; Sequence 18, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E
; APPLICANT: Gulick, Andrew M
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-297-431B-18
Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 27
US-08-297-431B-20
; Sequence 20, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E
; APPLICANT: Gulick, Andrew M
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-297-431B-20
Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131
```

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/297,431B
;; FILING DATE: August 26, 1994
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Reed, Janet E.
;; REGISTRATION NUMBER: 36252
;; REFERENCE/DOCKET NUMBER: WARE F039
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-972-8386
;; TELEFAX: 215-972-2292
;; INFORMATION FOR SEQ ID NO: 20:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 221 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-297-431B-20

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
| | | | |
Db 125 KDKARNR 131

RESULT 28
US-08-565-386-17
; Sequence 17, Application US/08555386
; Patent No. 5741697
; GENERAL INFORMATION:
; APPLICANT: Bavoll, Patrik M.
; APPLICANT: Hsia, Ru-ching
; TITLE OF INVENTION: BACTERIOPHAGE OF CHLAMYDIA PSITTACI
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Roches
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/565,386
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 176/60040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1636
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptidase
US-08-565-386-17

Query Match 0.6%; Score 7; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 849 NYIVNDS 855
| | | | |

Db 36 NYIVNDS 42

RESULT 29
US-08-576-626A-33
; Sequence 33, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:
; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Staver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; TITLE OF INVENTION: BIOSYNTHESIS GENES
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.O1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 322 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998194e
US-08-576-626A-33
Query Match 0.6%; Score 7; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 604 RQLITLL 610
| | | | |
Db 9 RQLITLL 15
RESULT 30
US-08-791-115B-27
; Sequence 27, Application US/08791115B
; Patent No. 6262242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershous, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
STREET: 555 Thirteenth Street, N.W., Suite 701-E
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 22204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/791.115B
APPLICATION NUMBER: US/08/791.115B
FILING DATE: 30-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Innen, Jeffrey L.
REGISTRATION NUMBER: 38,957
REFERENCE/DOCKET NUMBER: 2318-134.A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-683-6040
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-791-115B-27

Query Match 0.6%; Score 7; DB 4; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAEAL 645
DB 146 LKAEAL 152

RESULT 31
US-08-791-115B-1
; Sequence 1, Application US/08791115B
; Patent No. 6262242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershouse, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavtighian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791.115B
; FILING DATE: 30-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Innen, Jeffrey L.
; REGISTRATION NUMBER: 38,957
; REFERENCE/DOCKET NUMBER: 2318-134.A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-683-6040

TELEFAX: 202-683-7031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-791-115B-1
Query Match 0.6%; Score 7; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 639 LKAEAL 645
DB 146 LKAEAL 152

RESULT 32
US-08-791-115B-23
; Sequence 23, Application US/08791115B
; Patent No. 6262242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershouse, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavtighian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791.115B
; FILING DATE: 30-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Innen, Jeffrey L.
; REGISTRATION NUMBER: 38,957
; REFERENCE/DOCKET NUMBER: 2318-134.A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-683-6040
; TELEFAX: 202-683-7031
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-791-115B-23
Query Match 0.6%; Score 7; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 639 LKAEAL 645
DB 146 LKAEAL 152

RESULT 33
US-08-791-115B-25
; Sequence 25, Application US/08791115B


```
; Patent No. 6262242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershouse, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10023.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,115B
; FILING DATE: 30-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Innen, Jeffrey L.
; REGISTRATION NUMBER: 38,957
; REFERENCE/DOCKET NUMBER: 2318-134.A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-683-7031
; TELEFAX: 202-683-7031
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-791-115B-25

Query Match 0.6%; Score 7; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQAL 645
Db 146 LKAQAL 152
|||||

RESULT 34
US-08-791-115B-7
; Sequence 7, Application US/08791115B
; Patent No. 6262242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershouse, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10023.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,115B
; FILING DATE: 30-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Innen, Jeffrey L.
; REGISTRATION NUMBER: 38,957
; REFERENCE/DOCKET NUMBER: 2318-134.A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-683-7031
; TELEFAX: 202-683-7031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-791-115B-7

Query Match 0.6%; Score 7; DB 4; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQAL 645
Db 182 LKAQAL 188
|||||

RESULT 35
US-08-933-821-6
; Sequence 6, Application US/08933821
; Patent No. 5972338
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,821
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-933-821-6

Query Match 0.6%; Score 7; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRAIRKE 1030
|||||
```

Db 93 VRLRKE 99

RESULT 36
US-08-960-507-6
; Sequence 6, Application US/08960507
; Patent No. 6057435
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,507
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130p1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-960-507-6

Query Match 0.6%; Score 7; DB 3; Length 470;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1024 VRLRKE 1030

Db 93 VRLRKE 99

RESULT 37
US-09-136-828-6
; Sequence 6, Application US/09136828
; Patent No. 6350450
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,828
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130R1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-09-136-828-6

Query Match 0.6%; Score 7; DB 4; Length 470;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1024 VRLRKE 1030

Db 93 VRLRKE 99

RESULT 38
US-09-332-928A-6
; Sequence 6, Application US/09332928A
; Patent No. 6368833
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332,928A
; FILING DATE: 14-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/933,821
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-332-928A-6

Query Match 0.6%; Score 7; DB 4; Length 470;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030
 Db 93 VRLRKE 99

RESULT 39

US-09-136-801-6
 ; Sequence 6, Application US/09136801
 ; Patent No. 6413770
 ; GENERAL INFORMATION:
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin
 ; APPLICANT: Hillan, Kenneth
 ; APPLICANT: Botstein, David
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Roy, Margaret
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Schwall, Ralph
 ; TITLE OF INVENTION: Tie Ligand Homologues
 ; NUMBER OF SEQUENCES: 35
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA: US/09/136, 801
 ; APPLICATION NUMBER: US/09/136, 801

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.

REGISTRATION NUMBER: 33,055

REFERENCE/DOCKET NUMBER: P1130P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-3216

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 470 amino acids

TYPE: Amino Acid

TOPOLOGY: Linear

US-09-136-801-6

Query Match 0.6%; Score 7; DB 4; Length 470;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030
 Db 93 VRLRKE 99

RESULT 40

US-09-332-929-6
 ; Sequence 6, Application US/09332929
 ; Patent No. 6420542
 ; GENERAL INFORMATION:
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; TITLE OF INVENTION: Tie Ligands
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco

STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/332, 929
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/933, 821
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dreger, Ginger R.
 ; REGISTRATION NUMBER: 33,055
 ; REFERENCE/DOCKET NUMBER: P1130
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-3216
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 470 amino acids
 ; TYPE: Amino Acid
 ; TOPOLOGY: Linear
 ; US-09-332-929-6

Query Match 0.6%; Score 7; DB 4; Length 470;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030
 Db 93 VRLRKE 99

RESULT 41

US-08-968-563-15
 ; Sequence 15, Application US/08968563
 ; Patent No. 6013494
 ; GENERAL INFORMATION:
 ; APPLICANT: CHARLES E. NAKAMURA
 ; APPLICANT: ANTHONY A. GATENBY
 ; APPLICANT: AMY (KUANG-HUA) HSU
 ; APPLICANT: RICHARD D. LA REAU
 ; APPLICANT: SHARON L. HAYNIE
 ; APPLICANT: MARIA DIAZ-TORRES
 ; APPLICANT: DONALD E. TRIMBUR
 ; APPLICANT: GREGORY M. WHITED
 ; APPLICANT: VASANTHA NAGARAJAN
 ; APPLICANT: MARK S. PAYNE
 ; APPLICANT: STEPHEN K. PICATAGGIO
 ; APPLICANT: RAMESCH V. NAIR
 ; TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
 ; TITLE OF INVENTION: PRODUCTION OF 1,3-PROPANEDIOL
 ; NUMBER OF SEQUENCES: 49
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
 ; STREET: 1007 MARKET STREET
 ; CITY: WILMINGTON
 ; STATE: DELAWARE
 ; COUNTRY: U.S.A.
 ; ZIP: 19898
 ; ADDRESSEE: GENECOR INTERNATIONAL, INC.
 ; STREET: 4 CAMBRIDGE PLACE
 ; STREET: 1870 SOUTH WINTON ROAD
 ; CITY: ROCHESTER
 ; STATE: NEW YORK
 ; COUNTRY: U.S.A.
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.50 INCH DISKETTE

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,563
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,601
FILING DATE: NOVEMBER 13, 1996
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9982
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 501 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE: GLPD
US-08-968-563-15

Query Match 0.6%; Score 7; DB 3; Length 501;
Best Local Similarity 100.0%; Pred. No. 2.2e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 309 EHALEKL 315
Db 363 EHALEKL 369

RESULT 42
US-08-969-683A-15
Sequence 15, Application US/08969683A
Patent No. 6136576
GENERAL INFORMATION:
APPLICANT: GENENCOR INTERNATIONAL, INC.
TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
PRODUCTION OF 1,3 PROPANEDIOL
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 4 Cambridge Place
STREET: 1870 South Winton road
CITY: Rochester
STATE: NY
COUNTRY: U.S.A
ZIP: 14618
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/969,683A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/20873
FILING DATE: 13-NOV-1997
APPLICATION NUMBER: 60/030,601
FILING DATE: 13-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Graister, Dedra
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC 369-2
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-864-7620
TELEFAX: 650-845-6504
TELEX:

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 501 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE: GLPD
US-08-969-683A-15

Query Match 0.6%; Score 7; DB 4; Length 501;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 309 EHALEKL 315
Db 363 EHALEKL 369

RESULT 43
US-09-297-928-11
Sequence 11, Application US/09297928
Patent No. 6358716
GENERAL INFORMATION:
APPLICANT: BULTHUIS, BEN A.
GATENBY, ANTHONY A.
HAYNIE, SHARON L.
HSU, AMY K.
LAREAU, RICHARD D.
TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF
GLYCEROL BY RECOMBINANT
ORGANISMS

NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENENCOR INTERNATIONAL, INC.
STREET: 4 CAMBRIDGE PLACE
CITY: ROCHESTER
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 14618

COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/297,928
FILING DATE: 11-May-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/03602
FILING DATE: NOVEMBER 13, 1996
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9981-P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
TELEX: 6717325

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 501 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-09-297-928-11

Query Match 0.6%; Score 7; DB 4; Length 501;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 EHALEKL 315
| | | | |
DB 363 EHALEKL 369

RESULT 44

US-08-906-156A-12
; Sequence 12, Application US/08906156A
; Patent No. 6287854
; GENERAL INFORMATION:
; APPLICANT: SPURR, NIGEL K
; APPLICANT: GRAY, IAN C
; APPLICANT: STEWART, LORNA M
; TITLE OF INVENTION: DIAGNOSIS OF SUSCEPTIBILITY TO CANCER
; TITLE OF INVENTION: AND TREATMENT THEREOF
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,156A
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,655
; FILING DATE: 02-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,147
; FILING DATE: 13-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,840
; FILING DATE: 23-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/96GB/02588
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1090-14
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: translation of partial cDNA sequence

US-08-906-156A-12

Query Match 0.6%; Score 7; DB 4; Length 565;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQEAAL 645
| | | | |
DB 308 LKAQEAAL 314

RESULT 45

US-09-183-266A-49
; Sequence 49, Application US/09183266A
; Patent No. 6361954
; GENERAL INFORMATION:
; APPLICANT: Stillman, Bruce
; APPLICANT: Williams, R. Sanders
; APPLICANT: Mendez, Juan
; TITLE OF INVENTION: ANTIBODIES THERETO AND DIAGNOSTIC APPLICATIONS THEREOF
; TITLE OF INVENTION: DNA REPLICATION-REGULATING GENES,
; FILE REFERENCE: CSHL96-01A3
; CURRENT APPLICATION NUMBER: US/09/183,266A
; CURRENT FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: PCT/US97/07333
; PRIOR FILING DATE: 1997-05-02
; PRIOR APPLICATION NUMBER: 08/648,650
; PRIOR FILING DATE: 1996-05-15
; PRIOR APPLICATION NUMBER: 08/643,034
; PRIOR FILING DATE: 1996-05-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 577
; TYPE: PRT
; ORGANISM: S. cerevisiae
US-09-183-266A-49

Query Match 0.6%; Score 7; DB 4; Length 577;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 KLSLRKS 578
| | | | |
DB 157 KLSLRKS 163

RESULT 46

US-08-791-115B-6
; Sequence 6, Application US/08791115B
; Patent No. 6262242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershouse, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavtighian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,115B
; FILING DATE: 30-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 38,957
; REFERENCE/DOCKET NUMBER: 2318-134.A

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-683-6040
; TELEFAX: 202-683-7031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 645 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-791-115B-6

Query Match 0.6%; Score 7; DB 4; Length 645;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQAL 645
Db 388 LKAQAL 394
|||||

RESULT 47

; Sequence 5, Application US/08791115B
; Patent No. 626242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershouse, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavligian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,115B
; FILING DATE: 30-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 38,957
; REFERENCE/DOCKET NUMBER: 2318-134.A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-683-6040
; TELEFAX: 202-683-7031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 742 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-791-115B-5

Query Match 0.6%; Score 7; DB 4; Length 742;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQAL 645
Db 485 LKAQAL 491
|||||

RESULT 48

; TITLE OF INVENTION: MAMMALIAN REGULATOR OF
; NONSENSE-MEDIATED RNA DECAY

US-09-143-571-29
; Sequence 29, Application US/09143571
; Patent No. 633153
; GENERAL INFORMATION:
; APPLICANT: FISHEL, Richard
; APPLICANT: GRADIA, Scott
; APPLICANT: ACHARYA, Samir
; TITLE OF INVENTION: NUCLEOTIDE MODULATION OF DNA MISMATCH RECOGNITION
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR EFFECTING ADENINE
; FILE REFERENCE: 9855-601
; CURRENT APPLICATION NUMBER: US/09/143,571
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: 60/093,935
; EARLIER FILING DATE: 1998-07-23
; EARLIER APPLICATION NUMBER: 60/066,977
; EARLIER FILING DATE: 1997-11-28
; EARLIER APPLICATION NUMBER: 60/057,136
; EARLIER FILING DATE: 1997-08-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 834
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-143-571-29

Query Match 0.6%; Score 7; DB 4; Length 834;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 EFSDDLH 734
Db 366 EFSDDLH 372
|||||

RESULT 49

US-09-741-150-4
; Sequence 4, Application US/097411150
; Patent No. 6436689
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000968
; CURRENT APPLICATION NUMBER: US/09/741,150
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 884
; TYPE: PRT
; ORGANISM: Human
; US-09-741-150-4

Query Match 0.6%; Score 7; DB 4; Length 884;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 619 VLEQKQK 625
Db 193 VLEQKQK 199
|||||

RESULT 50

US-08-724-354D-22
; Sequence 22, Application US/08724354D
; Patent No. 5994119
; GENERAL INFORMATION:
; APPLICANT: Dietz, Harry C.
; TITLE OF INVENTION: MAMMALIAN REGULATOR OF
; NONSENSE-MEDIATED RNA DECAY

; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,354D
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/016,482
; FILING DATE: 29-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/090001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-678-5039
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 971 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-724-354D-22

Query Match 0.6%; Score 7; DB 2; Length 971;
Best Local Similarity 100.0%; Pred.No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 994 KTEAEIL 1000

Db 538 KTEAEIL 544

Search completed: November 6, 2002, 03:40:49
Job time : 93 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 23:19:32 : Search time 4531 Seconds
(without alignments)
13335.988 Million cell updates/sec

Title: US-09-782-874-1
Perfect score: 3731
Sequence: 1 GAATATCTTACTTACTT.....AGTTTCATCTTCTCTCTAAA 3731

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Listing first 50 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	575	15.4	659	13	BI203182
2	546	14.6	618	10	BE435332
3	525	14.1	525	10	AW933692
4	421	11.3	421	9	AI896465
5	398	10.7	570	10	BE433846
6	358	9.6	537	10	AW932918

7	341	9.1	597	10	AW649811
8	316	8.5	367	12	BF112386
9	195	5.2	725	12	BG594667
10	141	3.8	521	14	BQ111608
11	141	3.8	610	12	BE920043
12	121	3.2	434	14	BQ111609
13	110	2.9	403	12	BF052042
14	101	2.7	682	12	BG597875
15	98	2.6	773	14	BQ507474
16	95	2.5	443	10	AW623969
17	84	2.3	529	12	BG592782
18	81	2.2	393	12	BG889665
19	78	2.1	321	10	BE340733
20	71	1.9	550	9	AJ487433
21	57	1.5	340	14	BQ509665
22	24	0.6	541	14	BQ827639
23	23	0.6	348	17	BH258389
24	23	0.6	571	14	BQ488250
25	22	0.6	265	10	BB013867
26	22	0.6	308	9	AV045785
27	22	0.6	312	9	AV045898
28	22	0.6	332	9	AV210762
29	22	0.6	356	9	AV046870
30	22	0.6	580	10	AW644365
31	21	0.6	275	10	AV332932
32	21	0.6	282	9	AV207607
33	21	0.6	284	12	BG379160
34	21	0.6	323	10	BH446629
35	21	0.6	381	9	AV202914
36	21	0.6	426	10	BE581464
37	21	0.6	440	17	AQ498746
38	21	0.6	484	12	BG705217
39	21	0.6	500	13	BI693834
40	21	0.6	504	17	A2930706
41	21	0.6	521	13	BI038396
42	21	0.6	529	13	BI041199
43	21	0.6	630	17	AQ885038
44	21	0.6	637	17	A2866201
45	21	0.6	755	17	BH098015
46	21	0.6	765	17	BH608326
47	21	0.6	805	13	BG985828
48	21	0.6	808	17	BH558896
49	21	0.6	954	12	BF621359
50	21	0.6	1009	17	CNS0261X

ALIGNMENTS

RESULT 1
BI203182
LOCUS
DEFINITION
BI203182 cTOS Lycopersicon esculentum cDNA clone cTOS1C7 5' end similar to putative RNA-directed RNA polymerase, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 (bases 1 to 659)
AUTHORS
van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R., Ronning, C. and Tanksley, S.
TITLE
JOURNAL
COMMENT
Generation of ESTs from Tomato Suspension Cultures
Unpublished (2001)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.


```

FEATURES          Location/Qualifiers
  source          1..659
    /organism="Lycopersicon esculentum"
    /cultivar="TA496, E6203"
    /db_xref="taxon:4081"
    /clone="cFOS1C7"
    /clone_lib="cFOS"
    /tissue_type="suspension cultures"
    /lab_host="SOLR"
    /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Suspension cultures of L.esculentum E6203 were grown
in Murashige and Skoog based medium, supplemented with 1%
coconut milk (filter sterilized and added after
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
Fresh medium was added every 7 days, and cultures were
grown at 25 C, with 12hrs of light and continuous
shaking."
  BASE COUNT      210 a 107 c 153 g 189 t
  ORIGIN
    Query Match      15.4%; Score 575; DB 13; Length 659;
    Best Local Similarity 100.0%; Pred. No. 8.8e-294;
    Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 3060 CATATGATGCTGATGATGAGCTGATGATTTGAAGATTACATGACGAGCTTTTGACT 3119
  Db 1 CATATGATGCTGATGATGAGCTGATGATTTGAGATTACATGACGAGCTTTTGACT 50
  QY 3120 ACATAACTGATATGACAAAGCTGGGTATTTAATGAGCTACTATGGCATATAAAACAG 3179
  Db 61 ACATAACTGATATGACAAAGCTGGGTATTTAATGAGCTACTATGGCATATAAAACAG 120
  QY 3180 AGGCTGAATACTAGTGGTGGCATATGAGGATCAAAACTTTGACCGCAGAAAAG 3239
  Db 121 AGGCTGAATACTAGTGGTGGCATATGAGGATCAAAACTTTGACCGCAGAAAAG 180
  QY 3240 ATGCTGAGGCCATATGATGCTGAGGGCTTGAGGAAGGAGGCAAGAGCTGTTCA 3299
  Db 181 ATGCTGAGGCCATATGATGCTGAGGGCTTGAGGAAGGAGGCAAGAGCTGTTCA 240
  QY 3300 AGAGCGGTAATGATATGATGACATGTTACCAAGGCTTCGGCTTGGTACCACTACAT 3359
  Db 241 AGAGCGGTAATGATATGATGACATGTTACCAAGGCTTCGGCTTGGTACCACTACAT 300
  QY 3360 ATCATCTACATATGAGGCTGCTACATCAGGGTTGAAAGAGCTCATTTCAATGACT 3419
  Db 301 ATCATCTACATATGAGGCTGCTACATCAGGGTTGAAAGAGCTCATTTCAATGACT 360
  QY 3420 TTCCCTGGTGTGTTATGACACAGCTAATCCAGATTAAAGAGGACAAAGCAGCTAACAGGC 3479
  Db 361 TTCCCTGGTGTGTTATGACACAGCTAATCCAGATTAAAGAGGACAAAGCAGCTAACAGGC 420
  QY 3480 CAGTTCTCAACTGTCTCTCAGGGCTCAACTGAGTCACAGATTAGTGTGAATGAG 3539
  Db 421 CAGTTCTCAACTGTCTCTCAGGGCTCAACTGAGTCACAGATTAGTGTGAATGAG 480
  QY 3540 ATTCCAGTCGAGCGTTAAGCTGATATATATATATGATGATGAGGTTGATCATATAAGAAA 3599
  Db 481 ATTCCAGTCGAGCGTTAAGCTGATATATATATATGATGATGAGGTTGATCATATAAGAAA 540
  QY 3600 CTGTTATGATGTTGACTACCTTTTCTTTTAA 3634
  Db 541 CTGTTATGATGTTGACTACCTTTTCTTTTAA 575

  RESULT 2
  BE435332
  LOCUS
  DEFINITION
    ESP406410 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
  clone cLEG26C9, mRNA sequence.
  ACCESSION
    BE435332
  VERSION
    BE435332.1 GI:9433175
  KEYWORDS
    EST.

```

tomato.

SOURCE
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 618)

REFERENCE
AUTHORS
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
Location/Qualifiers
source
1..618
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG26C9"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmuvadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 196 a 115 c 149 g 157 t 1 others
ORIGIN

Query Match 14.6%; Score 546; DB 10; Length 618;
Best Local Similarity 100.0%; Pred. No. 2.3e-278;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2830 CAAAAAATTCGTGAGCTCTTTCAATTGACGTGGACTTCCAAAGACTGGTGTCCCGC 2889
Db 31 CAAAAAATTCGTGAGCTCTTTCAATTGACGTGGACTTCCAAAGACTGGTGTCCCGC 90
QY 2890 TGAATACCATCTCATGTCGCCCTTAAGAATACCCAGACTTCATGGATAGCCGACAA 2949
Db 91 TGAATACCATCTCATGTCGCCCTTAAGAATACCCAGACTTCATGGATAGCCGACAA 150
QY 2950 GACCAGCTATATCTCAGAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAGGACAAAGC 3009
Db 151 GACCAGCTATATCTCAGAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAGGACAAAGC 210
QY 3010 ACCTCAGGCTAGCTCTATCGGACCTTCACAAGAGATGTTCAAGGAGATCATATGATGC 3069
Db 211 ACCTCAGGCTAGCTCTATCGGACCTTCACAAGAGATGTTCAAGGAGATCATATGATGC 270
QY 3070 TGATATGGAAGTTGATGATTTGAAGATPACATTCAGCAAGCTTTTGACTACAAAATCTGA 3129
Db 271 TGATATGGAAGTTGATGATTTGAAGATPACATTCAGCAAGCTTTTGACTACAAAATCTGA 330
QY 3130 ATATGACAACAAGCTGGGTAAATTTAATGGACTACTATGGCATATAAAACAGAGGCTGAAAT 3189
Db 331 ATATGACAACAAGCTGGGTAAATTTAATGGACTACTATGGCATATAAAACAGAGGCTGAAAT 390
QY 3190 ACTTAGTGGGATATGAGGATCAAAAACCTTTTGACCGCAGAAAGATGCTGAGGC 3249
Db 391 ACTTAGTGGGATATGAGGATCAAAAACCTTTTGACCGCAGAAAGATGCTGAGGC 450
QY 3250 CATTAGTGTGCTGTGAGGCGCTTGAGGAAGGAGGAGAGAGCTGTTTCAAGAGGCGTAA 3309
Db 451 CATTAGTGTGCTGTGAGGCGCTTGAGGAAGGAGGAGAGAGAGCTGTTTCAAGAGGCGTAA 510

```
QY 3310 TGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTACATATCATCTAC 3369
|||||
Db 511 TGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTACATATCATCTAC 570
QY 3370 ATATTG 3375
|||||
Db 571 ATATTG 576

RESULT 3
AW933692
LOCUS
DEFINITION
EST359535 tomato fruit mature green, TAMU Lycopersicon esculentum
CDNA clone cLEF56C18 5', mRNA sequence.
AW933692
VERSION
AW933692.1 GI:8109093
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE
1. (bases 1 to 525)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
1..525
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF56C18"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT 151 a 111 c 121 g 142 t
ORIGIN
Query Match 14.1%; Score 525; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.3e-267;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTCATCCCGGTGATTCGTTTAAAGGCTGTAATGTTCCGAGCGCTGCA 2523
|||||
Db 1 TCCATGCTTCATCCCGGTGATTCGTTTAAAGGCTGTAATGTTCCGAGCGCTGCA 60
QY 2524 CCATGTTAGATGTTGTTATTCCTCCAGAAAGGAAAAAGACCCATCCCGAATGAATG 2583
|||||
Db 61 CCATGTTAGATGTTGTTATTCCTCCAGAAAGGAAAAAGACCCATCCCGAATGAATG 120
QY 2584 TTCTGGAGTATTTGGATGGGATATCTACTTTGTTCTGGGATCAAGACATGATCCC 2643
|||||
Db 121 TTCTGGAGTATTTGGATGGGATATCTACTTTGTTCTGGGATCAAGACATGATCCC 180
QY 2644 GCCAAGGCAAGTCCACCGATGAATATCTCCAGACCCAGACATACAGTTGGACCATGA 2703
|||||
Db 181 GCCAAGGCAAGTCCACCGATGAATATCTCCAGACCCAGACATACAGTTGGACCATGA 240
```

```
QY 2704 TGTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAATGACAGTTGGG 2763
|||||
Db 241 TGTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAATGACAGTTGGG 300
QY 2764 AATCATAGCAAAATGCCCATGTGCTATTTTCAGACAGAGAACCTGATATGCCATGAGTGA 2823
|||||
Db 301 AATCATAGCAAAATGCCCATGTGCTATTTTCAGACAGAGAACCTGATATGCCATGAGTGA 360
QY 2824 TCCATGCAAAAAAAGCTTGTGAGCTCTTTTCAATTGCAGTGGACTTTTCCAAAGACTGGTGT 2883
|||||
Db 361 TCCATGCAAAAAAAGCTTGTGAGCTCTTTTCAATTGCAGTGGACTTTTCCAAAGACTGGTGT 420
QY 2884 TCCCGCTGAAATACCATCTCAGTTGGCCCTTAAAGAAATACCCAGACTTCATGGATAAGCC 2943
|||||
Db 421 TCCCGCTGAAATACCATCTCAGTTGGCCCTTAAAGAAATACCCAGACTTCATGGATAAGCC 480
QY 2944 GGACAAGACCAAGCTATATCTCAGAAAGAGTATTGGAAGCTTTT 2988
|||||
Db 481 GGACAAGACCAAGCTATATCTCAGAAAGAGTATTGGAAGCTTTT 525

RESULT 4
AW96465
LOCUS
DEFINITION
EST265896 tomato callus, TAMU Lycopersicon esculentum CDNA clone
cLEC15G23, mRNA sequence.
AW96465
ACCESSION
AW96465.1 GI:5602355
VERSION
AW96465.1
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE
1. (bases 1 to 421)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
1..421
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC15G23"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato callus EST Library"
BASE COUNT 117 a 85 c 101 g 118 t
ORIGIN
Query Match 11.3%; Score 421; DB 9; Length 421;
Best Local Similarity 100.0%; Pred. No. 6.5e-212;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTCATCCCGGTGATTCGTTTAAAGGCTGTAATGTTCCGAGCGCTGCA 2523
|||||
```

```

Db 1 TCCATGCTTCATCCTCGTGATATTCCTGTTTAAAGGCTGTAATGTTTCGAGCGGTGCA 60
QY 2524 CCACATGGTAGATGTGTTGTATTCCTCCAGAAAGGAAAAAGACCTCATCCGAATG 2583
Db 61 CCACATGGTAGATGTGTTGTATTCCTCCAGAAAGGAAAAAGACCTCATCCGAATG 120
QY 2584 TTCGGGAGTGATTGGATGGGATATCTACTTTGTTGGTGGATCAAGACATGATCC 2643
Db 121 TTCGGGAGTGATTGGATGGGATATCTACTTTGTTGGTGGATCAAGACATGATCC 180
QY 2644 GCCAAGCAAGTCAGCCGATGGAATATCCTCCAGCACCAGCATACATGTCGACATGA 2703
Db 181 GCCAAGCAAGTCAGCCGATGGAATATCCTCCAGCACCAGCATACATGTCGACATGA 240
QY 2704 TGTCACAATTGAGGAAGTTGAAGACTTCAACCACTATATGTAATGACAGTTTGGG 2763
Db 241 TGTCACAATTGAGGAAGTTGAAGACTTCAACCACTATATGTAATGACAGTTTGGG 300
QY 2764 AATCATGCAAAATGCCCATGCTGTTATTCGAGACAGAACCTGATATGCCCATGAGTGA 2823
Db 301 AATCATGCAAAATGCCCATGCTGTTATTCGAGACAGAACCTGATATGCCCATGAGTGA 360
QY 2824 TCCATGCAAAATCTCTCAGCTCTTTTCAATTGAGTGGACTTCCAAAGACTGGTGT 2883
Db 361 TCCATGCAAAATCTCTCAGCTCTTTTCAATTGAGTGGACTTCCAAAGACTGGTGT 420
QY 2884 T 2884
Db 421 T 421

RESULT 5
LOCUS BE433646 570 bp mRNA linear EST 18-MAY-2001
DEFINITION EST400175 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone cLEG16J9, mRNA sequence.
ACCESSION BE433646
VERSION BE433646.1 GI:9431489
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 570)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Niernan,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
S.D.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
1..570
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG16J9"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSMCvAdapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end
of the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."

```

```

BASE COUNT 157 a 112 c 134 g 167 t
ORIGIN
Query Match 10.7%; Score 398; DB 10; Length 570;
Best Local Similarity 100.0%; Pred. NO. 1.2e-139;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2309 TTGATCAATCCAGAACCTTGGATATGTCAGGTGTTTGTTCAGTTTACTGGTCTGGA 2368
Db 173 TTGATCAATCCAGAACCTTGGATATGTCAGGTGTTTGTTCAGTTTACTGGTCTGGA 232
QY 2369 CATGGAGATTTTCTGACGATTTACATCCATTAAATACAGCAGATCCACCACAGTAAT 2428
Db 233 CATGGAGATTTTCTGACGATTTACATCCATTAAATACAGCAGATCCACCACAGTAAT 292
QY 2429 TTCAATCTGAAGGAATGTTGTTGCAAAAAATCCATGTTGCATCCTGGTGATAT 2488
Db 293 TTCAATCTGAAGGAATGTTGTTGCAAAAAATCCATGTTGCATCCTGGTGATAT 352
QY 2489 CGTGTTTTAAAGCTGTAAATGTTCCGAGCGCTGCACACATGTTAGATTGTTTATTC 2548
Db 353 CGTGTTTTAAAGCTGTAAATGTTCCGAGCGCTGCACACATGTTAGATTGTTTATTC 412
QY 2549 CCTCAGAAAGAAAGACCTCATCCGAATGATGTTCTGGGAGTGAATTTGGATGGGAT 2608
Db 413 CCTCAGAAAGAAAGACCTCATCCGAATGATGTTCTGGGAGTGAATTTGGATGGGAT 472
QY 2609 ATCTACTTTGTTCTGGGATCAAGACATGATCCGCCAAGCAAGTCCAGCCGATGAA 2668
Db 473 ATCTACTTTGTTCTGGGATCAAGACATGATCCGCCAAGCAAGTCCAGCCGATGAA 532
QY 2669 TATCCTCCAGACCCAGCATACAGTTGGACCATGATGT 2706
Db 533 TATCCTCCAGACCCAGCATACAGTTGGACCATGATGT 570

RESULT 6
LOCUS AW932918 537 bp mRNA linear EST 18-MAY-2001
DEFINITION EST358761 tomato fruit mature green, TAMU Lycopersicon esculentum
cDNA clone CLEF50H7 5', mRNA sequence.
ACCESSION AW932918
VERSION AW932918.1 GI:8108319
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 537)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
1..537
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEF50H7"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

```

Xhol: cLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"

BASE COUNT 177 a 112 c 118 g 130 t
ORIGIN

Query Match 9.6%; Score 358; DB 10; Length 537;
Best Local Similarity 99.8%; Pred. No. 2.2e-178;
Matches 408; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2635 CATGATCCGCGCAAGGCAAGTCCAGCCGATGGAATATCTCCAGACCCACCATACAGTT 2694

Db 1 CATGATCCGCGCAAGGCAAGTCCAGCCGATGGAATATCTCCAGACCCACCATACAGTT 60

QY 2695 GGACCATGATGTCACAATTTGAGGAAGTTGAAGAGTACTTCCCACTATATTTGGAATGA 2754

Db 61 GGACCATGATGTCACAATTTGAGGAAGTTGAAGAGTACTTCCCACTATATTTGGAATGA 120

QY 2755 CAGTTTGGGAATCATAGCAAAATGCCATGTCGTATTTTCGACACAGAGAACCTGATATGCC 2814

Db 121 CAGTTTGGGAATCATAGCAAAATGCCATGTCGTATTTTCGACACAGAGAACCTGATATGCC 180

QY 2815 CATGATGATCCATCCAAAACCTTCTGAGCTCTTTTCAATTCAGTGGACTTCCAAA 2874

Db 181 CATGATGATCCATCCAAAACCTTCTGAGCTCTTTTCAATTCAGTGGACTTCCAAA 240

QY 2875 GACTGGTGTCCCGCTGAAATACCATCTCAGTTGGCGCTTAAAGAAATACCCAGACTTCAT 2934

Db 241 GACTGGTGTCCCGCTGAAATACCATCTCAGTTGGCGCTTAAAGAAATACCCAGACTTCAT 300

QY 2935 GGATAGCGGGACAGACAGCATATATCTCAGAAAGAGTTATTTGGAAGCTTTTCAGGAA 2994

Db 301 GGATAGCGGGACAGACAGCATATATCTCAGAAAGAGTTATTTGGAAGCTTTTCAGGAA 360

QY 2995 AGTGAAGCAAGACACCTCAGGCTAGCTATCGACCTTCAAGA 3043

Db 361 AGTGAAGCAAGACACCTCAGGCTAGCTATCGACCTTCAAGA 409

RESULT 7
LOCUS AW649811 597 bp mRNA linear EST 18-MAY-2001
DEFINITION EST328265 tomato germinating seedlings, TAMU Lycopersicon esculentum cDNA clone cLEI9L3 5', mRNA sequence.

ACCESSION AW649811 GI:7411049

VERSION EST.

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 597)

AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley

,S.D.

Generation of ESTs from germinating tomato seed

Unpublished (2000)

Contact: CUGI

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

Location/Qualifiers

1..597

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/clone="cLEI9L3"

/clone_lib="tomato germinating seedlings, TAMU"

/tissue_type="whole seedlings"

/dev_stage="7 days post imbibition"

/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:

XhoI; 7 days post imbibition on water-agar. Mixed stage

whole germinating seedlings from seed coat emergence up

to two centimeters in seeds not showing obvious signs of

germination were discarded."

BASE COUNT 178 a 122 c 138 g 159 t

ORIGIN

Query Match 9.1%; Score 341; DB 10; Length 597;

Best Local Similarity 99.7%; Pred. No. 2.4e-169;

Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2572 TCCGAATGAATGTTCTGGAGTGTATTTGGATGGGGATATCTACTTTGTTGCTGGGATCA 2631

Db 149 TCCGAATGAATGTTCTGGAGTGTATTTGGATGGGGATATCTACTTTGTTGCTGGGATCA 208

QY 2632 AGACATATCCCGCAAGCAAGTCCAGCCGATGGATATCTCTCCAGCACCCAGCATACA 2691

Db 209 AGACATATCCCGCAAGCAAGTCCAGCCGATGGATATCTCTCCAGCACCCAGCATACA 268

QY 2692 GTTGGACCATGATGTCACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAA 2751

Db 269 GTTGGACCATGATGTCACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAA 328

QY 2752 TGACAGTTTGGGAATCATAGCAAAATGCCATGCCATGCGTATTTGCGAGACAGAACTGATAT 2811

Db 329 TGACAGTTTGGGAATCATAGCAAAATGCCATGCCATGCGTATTTGCGAGACAGAACTGATAT 388

QY 2812 GGCATCAGTGTCCCGCTGAAATACCATCTCAGTTGCGCCCTAAAGAAATACCCAGACTT 2871

Db 389 GGCATCAGTGTCCCGCTGAAATACCATCTCAGTTGCGCCCTAAAGAAATACCCAGACTT 448

QY 2872 AAAGACTGGTGTCCCGCTGAAATACCATCTCAGTTGCGCCCTAAAGAAATACCCAGACTT 2931

Db 449 AAAGACTGGTGTCCCGCTGAAATACCATCTCAGTTGCGCCCTAAAGAAATACCCAGACTT 508

QY 2932 CATGGATAAGCCGCGACAGACCCAGCTATATCT 2963

Db 509 CATGGATAAGCCGCGACAGACCCAGCTATATCT 540

RESULT 8 367 bp mRNA linear EST 18-MAY-2001

LOCUS BF112386 tomato breaker fruit Lycopersicon esculentum cDNA clone

DEFINITION EST439976 5' sequence, mRNA sequence.

ACCESSION BF112386

VERSION BF112386

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 367)

AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley

,S.D.

Generation of ESTs from tomato fruit tissue, breaker stage

Unpublished (2000)

Contact: CUGI

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

Location/Qualifiers

1..367

```

/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG41011"
/clone_lib="tomato breaker fruit"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/notes="Vector: pBluescriptSKmCudapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT      102 a  75 c  89 g  101 t
ORIGIN

```

```

Query Match      8.5%; Score 316; DB 12; Length 367;
Best Local Similarity 99.7%; Pred. No. 4.4e-156; Indels 0; Gaps 0;
Matches 366; Conservative 0; Mismatches 1;
QY 2464 TCCATGCTTGCATCCTGGTGATATTCGTTTAAAGGCTGTAATGTCGAGCGTGCA 2523
DB 1 TCCATGCTTGCATCCTGGTGATATTCGTTTAAAGGCTGTAATGTCGAGCGTGCA 60
QY 2524 CCATGGTAGATTGTTGTAATTCCTCAGAAAGGAAAGACCCTCATCCGATGATG 2583
DB 61 CCATGGTAGATTGTTGTAATTCCTCAGAAAGGAAAGACCCTCATCCGATGATG 120
QY 2584 TTCTGGGAGTAGTTGGATGGGATATCTACTTTGTTGGTCAAGACATGATCCC 2643
DB 121 TTCGGGAGTAGTTGGATGGGATATCTACTTTGTTGGTCAAGACATGATCCC 180
QY 2644 GCCAAGCAATCCAGCGATGGAATATCCTCCAGCAGCCAGCATACAGTTGNCCATGA 2703
DB 181 GCCAAGCAATCCAGCGATGGAATATCCTCCAGCAGCCAGCATACAGTTGNCCATGA 240
QY 2704 TGTCACAATTGAGGAAGTTGAAGAGTACTTCCACCAACTATATTGTGAATGACAGTTGGG 2763
DB 241 TGTCACAATTGAGGAAGTTGAAGAGTACTTCCACCAACTATATTGTGAATGACAGTTGGG 300
QY 2764 AATCATAGCAATGCCATGTCGATTGTCAGACAGAGACCTGATATGGCCATGATGA 2823
DB 301 AATCATAGCAATGCCATGTCGATTGTCAGACAGAGACCTGATATGGCCATGATGA 360
QY 2824 TCCATGC 2830
DB 361 TCCATGC 367

```

```

RESULT 9
BG594667
LOCUS      BG594667      725 bp      mRNA      linear      EST 12-APR-2001
DEFINITION EST493345 cSTS Solanum tuberosum cDNA clone cSTS8119 5' sequence,
mRNA sequence.
ACCESSION  BG594667
VERSION     BG594667.1  GI:13612807
KEYWORDS   EST.
SOURCE      potato.
ORGANISM   Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 725)
AUTHORS   van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,
Bougri,O., Buell,C.R., Ronning,C., Tanksley,S., and Baker,B.
TITLE      Generations of ESTs from sprouting potato eyes
JOURNAL    Unpublished (2000)
COMMENT    Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com

```

FEATURES

```

Seq primer: M13F-R.
Location/Qualifiers
1..725
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTS8119"
/clone_lib="cSTS"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."
BASE COUNT      228 a  140 c  170 g  187 t
ORIGIN

```

```

Query Match      5.2%; Score 195; DB 12; Length 725;
Best Local Similarity 99.6%; Pred. No. 1.1e-91; Indels 0; Gaps 0;
Matches 245; Conservative 0; Mismatches 1;
QY 2699 CATGATGTCACAAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAATGACAGT 2758
DB 212 CATGATGTCACAAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAATGACAGT 271
QY 2759 TTGGGAATCATAGCAATGCCATGTCGATTGGCAGACAGAGAACCTGATATGCCATG 2818
DB 272 TTGGGAATCATAGCAATGCCATGTCGATTGGCAGACAGAGAACCTGATATGCCATG 331
QY 2819 AGTGATCCATGCAAAAACCTTGTGAGCTCTTTTCAATTGCAAGTGGACTTCCAAAGACT 2878
DB 332 AGTGATCCATGCAAAAACCTTGTGAGCTCTTTTCAATTGCAAGTGGACTTCCAAAGACT 391
QY 2879 GGTGTTCCCGCTGAATACCATCTCAGTTGCCCTTAAGATATCCACACTTCATGGAT 2938
DB 392 GGTGTTCCCGCTGAATACCATCTCAGTTGCCCTTAAGATATCCACACTTCATGGAT 451
QY 2939 AAGCCG 2944
DB 452 AAGCCG 457

```

```

RESULT 10
BG111608
LOCUS      BG111608      521 bp      mRNA      linear      EST 22-JUL-2002
DEFINITION EST597184 mixed potato tissues Solanum tuberosum cDNA clone STMCB87
5' end mRNA sequence.
ACCESSION  BG111608
VERSION     BG111608.2  GI:21915035
KEYWORDS   EST.
SOURCE      potato.
ORGANISM   Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 521)
AUTHORS   Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karamecheva,S.A.
TITLE      Generation of a set of potato cDNA clones for microarray analyses
JOURNAL    Unpublished (2002)
COMMENT    On Apr 17, 2002 this sequence version replaced gi:20163570.
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato@tigr.org
This clone is available through the Research Genetics, contact the
Research Genetics for further information 1-800-711-6195 or
cdna@resgen.com
Seq primer: T3.

```

FEATURES	Location/Qualifiers	Leaflets and petioles were isolated from 8 week old greenhouse grown plants. The plants were watered and fertilized freely. The tissue was immediately frozen in liquid nitrogen.
source	1. .521	
	/organism="Solanum tuberosum"	
	/cultivar="Kennebec or Binjte"	
	/db_xref="taxon:4113"	
	/clone="STMCB87"	
	/clone_lib="mixed potato tissues"	
	/tissue_type="mixed potato tissues"	
	/lab_host="SOLR"	
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes 'tubers, or roots."	
BASE COUNT	155 a 106 c 121 g 139 t	
ORIGIN		
Query Match	3.8%; Score 141; DB 14; Length 521;	
Best Local Similarity	99.5%; Pred. No. 5.4e-63;	
Matches 191; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY 2753	GACAGTTGGGATCATAGCAATGCCATGCTGATTGTCAGACAGACAGACCTGATG 2812	
DB 274	GACAGTTGGGATCATAGCAATGCCATGCTGATTGTCAGACAGACAGACCTGATG 333	
QY 2813	GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGAGTGAGCTTTCCA 2872	
DB 334	GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGAGTGAGCTTTCCA 393	
QY 2873	AGACTGGTGTCCCGCTGAATACATCTCAGTTGGCCCTAAAGAAATACCCAGACTTC 2932	
DB 394	AGACTGGTGTCCCGCTGAATACATCTCAGTTGGCCCTAAAGAAATACCCAGACTTC 453	
QY 2933	ATGGATAAGCCG 2944	
DB 454	ATGGATAAGCCG 465	
RESULT 11		
BE920043		
LOCUS	610 bp mRNA linear EST 02-OCT-2000	
DEFINITION	EST423812 potato leaves and petioles Solanum tuberosum cDNA clone	
ACCESSION	BE920043	
VERSION	BE920043.1 GI:10446119	
KEYWORDS	potato.	
SOURCE	Solanum tuberosum	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.	
REFERENCE	1 (bases 1 to 610)	
AUTHORS	van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J., Otterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.	
TITLE	Generation of ESTs from potato leaves and petioles	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Cathy Ronning The Institute for Genomic Research For clone request: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com.	
FEATURES	Location/Qualifiers	
source	1. .610	
	/organism="Solanum tuberosum"	
	/cultivar="Kennebec"	
	/db_xref="taxon:4113"	
	/clone="cSPB3D19"	
	/clone_lib="potato leaves and petioles"	
	/tissue_type="leaflets and petioles"	
	/dev_stage="8 weeks old plants"	
	/lab_host="SOLR"	
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; tissue was supplied by Dr. Fry (Cornell University).	
BASE COUNT	129 a 95 c 73 g 137 t	
ORIGIN		
Query Match	3.2%; Score 121; DB44; Length 434;	
Best Local Similarity	99.5%; Pred. No. 5.5e-63;	
Matches 191; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY 2753	GACAGTTGGGATCATAGCAATGCCATGCTGATTGTCAGACAGACAGACCTGATG 2812	
DB 274	GACAGTTGGGATCATAGCAATGCCATGCTGATTGTCAGACAGACAGACCTGATG 333	
QY 2813	GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGAGTGAGCTTTCCA 2872	
DB 334	GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGAGTGAGCTTTCCA 393	
QY 2873	AGACTGGTGTCCCGCTGAATACATCTCAGTTGGCCCTAAAGAAATACCCAGACTTC 2932	
DB 394	AGACTGGTGTCCCGCTGAATACATCTCAGTTGGCCCTAAAGAAATACCCAGACTTC 453	
QY 2933	ATGGATAAGCCG 2944	
DB 454	ATGGATAAGCCG 465	
RESULT 12		
BE920043		
LOCUS	434 bp mRNA linear EST 17-APR-2002	
DEFINITION	EST597185 mixed potato tissues Solanum tuberosum cDNA clone STMCB87	
ACCESSION	BE920043	
VERSION	BE920043.1 GI:20163571	
KEYWORDS	potato.	
SOURCE	Solanum tuberosum	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.	
REFERENCE	1 (bases 1 to 434)	
AUTHORS	Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and Karamycheva,S.A.	
TITLE	Generation of a set of potato cDNA clones for microarray analyses	
JOURNAL	Unpublished (2002)	
COMMENT	Contact: Robin Buell The Institute for Genomic Research 9712 Medical Center Dr, Rockville, MD 20850, USA Email: potato@tigr.org This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cdna@resgen.com	
FEATURES	Location/Qualifiers	
source	1. .434	
	/organism="Solanum tuberosum"	
	/cultivar="Kennebec or Binjte"	
	/db_xref="taxon:4113"	
	/clone="STMCB87"	
	/clone_lib="mixed potato tissues"	
	/tissue_type="mixed potato tissues"	
	/lab_host="SOLR"	
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes 'tubers, or roots."	
BASE COUNT	129 a 95 c 73 g 137 t	
ORIGIN		
Query Match	3.2%; Score 121; DB44; Length 434;	
Best Local Similarity	99.5%; Pred. No. 5.5e-63;	
Matches 191; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY 2753	GACAGTTGGGATCATAGCAATGCCATGCTGATTGTCAGACAGACAGACCTGATG 2812	
DB 274	GACAGTTGGGATCATAGCAATGCCATGCTGATTGTCAGACAGACAGACCTGATG 333	
QY 2813	GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGAGTGAGCTTTCCA 2872	
DB 334	GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTGAGTGAGCTTTCCA 393	
QY 2873	AGACTGGTGTCCCGCTGAATACATCTCAGTTGGCCCTAAAGAAATACCCAGACTTC 2932	
DB 394	AGACTGGTGTCCCGCTGAATACATCTCAGTTGGCCCTAAAGAAATACCCAGACTTC 453	
QY 2933	ATGGATAAGCCG 2944	
DB 454	ATGGATAAGCCG 465	

Best Local Similarity 99.4%; Pred. No. 2.3e-52;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3291 CTTGGTTCAGAGCGCTATGATATAGATGACATGTTACCAAGCGCTTCGGCTGGTACC 3350
|||||
Db 419 CTGGTTCAGAGCGCTATGATATAGATGACATGTTACCAAGCGCTTCGGCTGGTACC 360
|||||
QY 3351 ACCTTACATATCATCTACATATTTGGGTGCTCAATCAGGCGTTGAAAAGAGCTCAT 3410
|||||
Db 359 ACCTTACATATCATCTACATATTTGGGTGCTCAATCAGGCGTTGAAAAGAGATCAT 300
|||||
QY 3411 TCATTAGCTTCCCTGGTGTGTTAAGACCACTAATCCAGATTAAGAAGA 3462
|||||
Db 299 TCATTAGCTTCCCTGGTGTGTTAAGACCACTAATCCAGATTAAGAAGA 248
|||||

RESULT 13
BF052042
LOCUS
DEFINITION
EST437289 tomato developing/immature green fruit Lycopersicon
esculentum cDNA clone cLEM25110 5' sequence, mRNA sequence.
ACCESSION
BF052042
VERSION
BF052042.1 GI:10805938
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE
AUTHORS
Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,
Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Ronning, C.M.,
Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley
S.D.

TITLE
JOURNAL
COMMENT
Generation of ESTs from tomato fruit tissue, immature green
Unpublished (2000)

CONTACT: CUGI
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.
Location/Qualifiers

FEATURES
Source
1..403
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEM25110"
/clone_lib="tomato developing/immature green fruit"
/tissue_type="fruit"
/dev_stage="immature green (5-35 days post-anthesis)"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCudapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and
harvested at 7 day intervals through 35 dpa. Equal masses
of tissue from each stage were combined (including seeds
and locules) prior to mRNA isolation."
BASE COUNT
128 a 68 c 97 g 110 t

Query Match 2.9%; Score 110; DB 12; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.6e-46;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3281 GAGGCAAGAGCGTGGTTCAGAGCGGTAATGATATAGATGACATGTTACCAAGGCTTCG 3340
|||||

Db 222 GAGGCAAGAGCGTGGTTCAGAGCGGTAATGATATAGATGACATGTTACCAAGGCTTCG 281
|||||

QY 3341 GCTTGGTACCAGGTACATATATCCTACATATTTGGGTTGCTACAATCA 3390
|||||

Db 282 GCTTGGTACCAGGTACATATATCCTACATATTTGGGTTGCTACAATCA 331
|||||

RESULT 14
BG597875

LOCUS
DEFINITION
EST496553 cDNAs Solanum tuberosum cDNA clone cSTS19C12 5' sequence,
mRNA sequence.
ACCESSION
BG597875
VERSION
BG597875.1 GI:13616015
KEYWORDS
EST.
SOURCE
potato.
ORGANISM
Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE
AUTHORS
1 (bases 1 to 682)
van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chieningo, A.,
Bougr, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.

Generations of ESTs from sprouting potato eyes
Unpublished (2000)

CONTACT: Cathy Ronning

The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.

FEATURES
Location/Qualifiers
1..682

/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTS19C12"
/clone_lib="cSTS"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."
BASE COUNT
185 a 108 c 154 g 235 t

ORIGIN

Query Match 2.7%; Score 101; DB 12; Length 682;
Best Local Similarity 100.0%; Pred. No. 1.1e-41;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 TATAAATTCAGCTTTCATATGAAATATATGCGAGGTGTGCTCCATCGTCATATGGT 673
|||||

Db 540 TATAAATTCAGCTTTCATATGAAATATATGCGAGGTGTGCTCCATCGTCATATGGT 599
|||||

QY 674 CAAAATGCTCAGTTCTCCTCATACAGTTATTTGGTGCTCC 714
|||||

Db 600 CAAAATGCTCAGTTCTCCTCATACAGTTATTTGGTGCTCC 640
|||||

RESULT 15
BQ507474

LOCUS
DEFINITION
EST614889 Generation of a set of potato cDNA clones for microarray
analyses mixed potato tissues Solanum tuberosum cDNA clone STMG059
5' end, mRNA sequence.
ACCESSION
BQ507474
VERSION
BQ507474.2 GI:21923304
KEYWORDS
EST.
SOURCE
potato.
ORGANISM
Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE
AUTHORS
1 (bases 1 to 773)
Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,
Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and
Karamycheva, S.A.

TITLE
JOURNAL
COMMENT
Generation of a set of potato cDNA clones for microarray analyses
On Jun 10, 2002 this sequence version replaced gi:21366343.
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato@tigr.org
This clone is available through the Research Genetics, contact the
Research Genetics for further information 1-800-711-6195 or
cdna@resgen.com
Seq primer: T3

FEATURES

source
1..773
/organism="Solanum tuberosum"
/cultivar="Kennebec Or Binite"
/db_xref="taxon:4113"
/clone="STM059"
/clone_lib="Generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
tubers, or roots."
BASE COUNT 205 a 125 c 181 g 262 t
ORIGIN

Query Match 2.6%; Score 98; DB 14; Length 773;
Best Local Similarity 100.0%; Pred. No. 4.2e-40;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 644 TGGCAGGTGCTCCATCGTCATATGGTCAAAATGCTCAGTTCTCTATACAGTTA 703
|||||
Db 577 TGGCAGGTGCTCCATCGTCATATGGTCAAAATGCTCAGTTCTCTATACAGTTA 636
|||||
QY 704 TTGTGTCTCCCGGATCTATAAGAGACTTGAAGAACTC 741
|||||
Db 637 TTGTGTCTCCCGGATCTATAAGAGACTTGAAGAACTC 674
|||||

RESULT 16
AW623969
LOCUS
DEFINITION
EST321914 tomato flower buds 3-8 mm, Cornell University
Lycopersicon esculentum cDNA clone cTOB13P9 5', mRNA sequence.
ACCESSION
AW623969
VERSION
AW623969.1 GI:7336996
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum

REFERENCE
AUTHORS
van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang
F., Hansen, T., Craven, M.B., Bowman, C.L., Ronning, C.M., Nierman, W.,
Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.
Generation of ESTs from tomato flower tissue, 3-8 mm buds
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source
Location/Qualifiers
1..443
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"

/clone="cTOB13P9"
/clone_lib="tomato flower buds 3-8 mm, Cornell University"
/tissue_type="flower"
/dev_stage="3-8mm buds"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA496). They
were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."
BASE COUNT 135 a 101 c 98 g 109 t
ORIGIN

Query Match 2.5%; Score 95; DB 10; Length 443;
Best Local Similarity 99.3%; Pred. No. 1.5e-38;
Matches 145; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2955 GCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAGCAAAAGCACCTC 3014
|||||
Db 293 GCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAGCAAAAGCACCTC 352
|||||
QY 3015 AGGCTAGCTCTATCGGCACCTTCACAAAGAGATGTTGCAAGGAGATCATATATGCTGTGATA 3074
|||||
Db 353 ATGCTAGCTCTATCGGCACCTTCACAAAGAGATGTTGCAAGGAGATCATATATGCTGTGATA 412
|||||
QY 3075 TGAAGTTGATGGATTTGAAGATTAC 3100
|||||
Db 413 TGAAGTTGATGGATTTGAAGATTAC 438
|||||

RESULT 17
BG592782
LOCUS
DEFINITION
EST491460 cSTS Solanum tuberosum cDNA clone cSTS2K4 5' sequence,
mRNA sequence.
ACCESSION
BG592782
VERSION
BG592782.1 GI:13610922
KEYWORDS
EST.
SOURCE
potato.
ORGANISM
Solanum tuberosum

REFERENCE
AUTHORS
van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chienming, A.,
Bougr, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.
Generations of ESTs from sprouting potato eyes
Unpublished (2000)
Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.

FEATURES
source
Location/Qualifiers
1..529
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTS2K4"
/clone_lib="cSTS"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."

BASE COUNT 168 a 103 c 120 g 138 t
ORIGIN
Query Match 2.3%; Score 84; DB 12; Length 529;
Best Local Similarity 99.3%; Pred. No. 1.1e-32;


```

Matches 134; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2715 AGGAAGTTGAAGTACTTCCACCAACTATATTGTGAATGACAGTTGGGAATCATAGCAA 2774
Db 25 AGGAAGTTGAAGTACTTCCACCAACTATATTGTGAATGACAGTTGGGAATCATATCAA 84
Qy 2775 ATGCCCATGCGTATTTCGACAGACAGAACTGATATGGCCATGAGTGATCCATGCAAAA 2834
Db 85 ATGCCCATGCGTATTTCGACAGACAGAACTGATATGGCCATGAGTGATCCATGCAAAA 144
Qy 2835 AACTTGCTGAGCTCT 2849
Db 145 AACTTGCTGAGCTCT 159

RESULT 18
BG889665 393 bp mRNA linear EST 30-MAY-2001
LOCUS EST515516 cSTD Solanum tuberosum cDNA clone cSTD15C7 5' sequence,
DEFINITION mRNA sequence.
ACCESSION BG889665
VERSION BG889665
KEYWORDS EST.
SOURCE BG889665.1 GI:14266751
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 393)
van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chieningo,A.,
Bougr,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
Generations of ESTs from dormant potato tubers
Unpublished (2001)
Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
FEATURES
source
1..393
Location/Qualifiers
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD15C7"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; This library targets genes expressed in dormant
tubers. This library was made from sections of dormant
tuber, avoiding the buds and epidermis. Tubers were stored
for one month post-harvest at 40C. The tuber was peeled,
well away from the surface. Then it was chopped into 1-2
mm cubes and immediately frozen in liquid nitrogen. This
library is noted as P4 in Tanksley lab notebooks."
BASE COUNT 114 a 60 c 96 g 123 t
ORIGIN
Query Match 2.2%; Score 81; DB 12; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.2e-31;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 270 ATACAGGATATGGAACTGATATGTCATGGAGGTTAAACAGTCCAAAGGAGGATCTAGAG 329
Db 152 ATACAGGATATGGAACTGATATGTCATGGAGGTTAAACAGTCCAAAGGAGGATCTAGAG 211
Qy 330 CATTGCCAAAGTTCAATTG 350
Db 212 CATTGCCAAAGTTCAATTG 232

RESULT 19
BG889665 321 bp mRNA linear EST 17-JUL-2000
LOCUS EST344811 potato stolon, Cornell University Solanum tuberosum cDNA
DEFINITION clone cSTA12M12 similar to RNA-directed RNA polymerase
(Lycopersicon esculentum) PIR[T30819][T30819 RNA-direc, mRNA
sequence.
ACCESSION BE340733
VERSION BE340733
KEYWORDS EST.
SOURCE BE340733.1 GI:9250264
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 321)
van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
Hoit,I.E., Liang,F., Hansen,T.S., Utterback,T., Bowman,C.L., Doan
B., Bougr,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker
,B.
Generation of ESTs from potato swelling stolons
Unpublished (1999)
Contact: Research Genetics, Libraries Division
Tel: 1-800-711-6195
Email: cdna@resgen.com
5 prime sequence.
FEATURES
source
1..321
Location/Qualifiers
/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTA12M12"
/clone_lib="potato stolon, Cornell University"
/tissue_type="axillary buds of stem explants, swelling
stolons"
/dev_stage="1 to 3 days"
/lab_host="SOLR"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; RNA was supplied by Christian Bachem & Beatrix
Horvath(Laboratory of Plant Breeding, Dept. of Plant
Sciences, Wageningen University, The Netherlands). Total
RNA was isolated from developing axillary buds of potato
nodal stem cuttings cultured on medium for the
introduction of tuber formation as described in Bachem et
al. (Plant Journal 1996). Tissue samples were taken of
stages corresponding to growing stolons and the early
stages of tuber formation."
BASE COUNT 87 a 51 c 80 g 103 t
ORIGIN
Query Match 2.1%; Score 78; DB 10; Length 321;
Best Local Similarity 99.2%; Pred. No. 1.6e-29;
Matches 128; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2243 AAGTTGCTCGATTGCGGACTAGATCAAGAATATTTATTCCAAATGGAAGACAATGATG 2302
Db 172 AAGTTGCTCGATTGCGGACTAGATCAAGAATATTTATTCCAAATGGAAGACAATGATG 231
Qy 2303 GGATGTTGGATGAATCCAGAACCTTGGATATGGTCAGGTGTTTGTTCAGTTACTGCT 2362
Db 232 GGATGTTGGATGAATCCAGAACCTTGGATATGGTCAGGTGTTTGTTCAGTTACTGCT 291
Qy 2363 GCTGGACAT 2371
Db 292 GCTGGACAT 300

RESULT 20
AJ487433 550 bp mRNA linear EST 30-JUL-2002
LOCUS AJ487433 Solanum tuberosum cv. Provita Solanum tuberosum cDNA clone
DEFINITION Pf8, mRNA sequence.
ACCESSION AJ487433
VERSION AJ487433.1 GI:22022186

```

KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 550)
AUTHORS Gebhardt, C., Walkemeier, B., Henselewski, H., Barakat, A., Delseny, M. and Stueber, K.

TITLE Comparative mapping between potato (Solanum tuberosum) and Arabidopsis thaliana reveals structurally conserved domains and ancient duplications in the potato genome

JOURNAL Unpublished (2002)
COMMENT Contact: Gebhardt C
MPI for Breeding Research
Carl-von-Linne Weg 10, Cologne, D-50829, Germany.

FEATURES
source
1..550
/organism="Solanum tuberosum"
/cultivar="Provita"
/db_xref="taxon:4113"
/map="v"
/clone="P3f8"
/clone_lib="Solanum tuberosum cv. Provita"

BASE COUNT 182 a 91 c 125 g 152 t
ORIGIN

Query Match 1.9%; Score 71; DB 9; Length 550;
Best Local Similarity 99.2%; Pred. No. 9.2e-26;
Matches 121; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3341 GCTTGGTACCGGTACATATCATCTACATATTGGGGTCTCAATCAGGGTTGAAA 3400
|||||
Db 222 GCTTGGTACCGGTACATATCATCTACATATTGGGGTCTCAATCAGGGTTGAAA 281
|||||

QY 3401 AGAGTCATTTTCATTTAGCTTCCCTGGTGTGTATGACCGAGCTAATCCAGATTAAAG 3460
|||||
Db 282 AGAGTCATTTTCATTTAGCTTCCCTGGTGTGTATGACCGAGCTAATCCAGATTAAAG 341
|||||

QY 3461 GA 3462
||
Db 342 GA 343

RESULT 21
BQ509665/c
LOCUS
DEFINITION EST617080 Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues Solanum tuberosum cDNA clone STMHE37 3', end, mRNA sequence.

ACCESSION BQ509665
VERSION BQ509665.2 GI:21925399
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 340)
AUTHORS Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C., Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and Karamycheva, S.A.

TITLE Generation of a set of potato cDNA clones for microarray analyses

JOURNAL Unpublished (2002)
COMMENT On Jun 10, 2002 this sequence version replaced gi:21368534.
Other_ESTs: EST617079
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato@tigr.org
This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or

cdna@resgen.com
Seq primer: T7
Location/Qualifiers
1..340
/organism="Solanum tuberosum"
/cultivar="Kennebec or Binjfe"
/db_xref="taxon:4113"
/clone="STMHE37"
/clone_lib="Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes, tubers, or roots."

BASE COUNT 107 a 53 c 55 g 115 t
ORIGIN

Query Match 1.5%; Score 57; DB 14; Length 340;
Best Local Similarity 100.0%; Pred. No. 2.4e-18;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3406 TCATTTCAATTAGCTTCCCTGGTGTGTATGACCGAGCTAATCCAGATTAAAGGA 3462
|||||
Db 315 TCATTTCAATTAGCTTCCCTGGTGTGTATGACCGAGCTAATCCAGATTAAAGGA 259
|||||

RESULT 22
BQ827639
LOCUS
DEFINITION BQ827639.1 GI:221333265
541 bp mRNA linear EST 07-AUG-2002
9058f03.y1 Moss EST library PPAS Physcomitrella patens cDNA clone
PEP_SOURCE_ID:PPAS041205 5' similar to TR:Q9SG02 Q9SG02 PUTATIVE
RNA-DIRECTED RNA POLYMERASE. ;, mRNA sequence.

ACCESSION BQ827639
VERSION BQ827639.1 GI:221333265
KEYWORDS EST.
SOURCE Physcomitrella patens.
ORGANISM Physcomitrella patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
1 (bases 1 to 541)
REFERENCE Quatrano, R., Bashardes, S., Cove, D., Cumming, A., Knight, C., Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R. and Wilson, R.
AUTHORS Quatrano, R. and Moss EST Project
Leeds/Wash U Moss EST Project
Unpublished (1999)
Contact: Ralph Quatrano
Leeds/Wash U Moss EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Libraries were constructed by Dr. Stavros Bashardes as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
Seq primer: -40RP from Gibco
High quality sequence stop: 447.

FEATURES
source
1..541
/organism="Physcomitrella patens"
/db_xref="taxon:3218"
/clone="PEP_SOURCE_ID:PPAS041205"
/clone_lib="Moss EST library PPAS"
/dev_stage="protonemata, 7day old ABA-treated"
/lab_host="E.coli DH10b"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

ECORI; 7-day-old pronemal tissue was incubated 10-4M ABA for 16 hours before RNA isolation. cDNA was synthesised from Amersham's cDNA Synthesis plus kit and ligated with EcoRI-NotI linkers for cloning in the EcoRI site of Lampda ZapII (Stratagene). After packaging, the library was propagated in E. coli XL-I Blue cells and amplified. The library was excised by mass excision in SOLR cells and ampicillin resistant transformants selected. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered and used to transform DH10b cells by electroporation. Clones corresponding to abundant transcripts were identified by colony hybridization using a cDNA probe derived from untreated pronemal tissue, and eliminated from the library, by rearraying."

BASE COUNT 133 a 113 c 141 g 154 t
 ORIGIN
 Query Match 0.6%; Score 24; DB 14; Length 541;
 Best Local Similarity 100.0%; Pred. No. 0.91;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2458 AAAAAATCCATGCTTCATCCTGG 2481
 |||||
 Db 349 AAAAAATCCATGCTTCATCCTGG 372

RESULT 23
 BH258389
 LOCUS
 DEFINITION BH258389 348 bp DNA linear GSS 30-NOV-2001
 CH230-112K20-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
 CH230-112K20, DNA sequence.
 ACCESSION BH258389.1 GI:17160712
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 348)
 AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
 A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
 Jong,P. and Fraser,C.M.
 TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment
 JOURNAL Unpublished (1999)
 COMMENT Other GSSs: CH230-112K20.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library
 availability, please contact Pieter de Jong (pjejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/orering_information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 112 row: K column: 20
 Seq primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1. .348
 /organism="Rattus norvegicus"
 /strain="BN/SSNHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-112K20"
 /clone_lib="CHORI-230 Segment 1"
 /sex="Female"
 /cell_type="Brain"
 /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;

FEATURES

source
 1. .348
 /organism="Rattus norvegicus"
 /strain="BN/SSNHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-112K20"
 /clone_lib="CHORI-230 Segment 1"
 /sex="Female"
 /cell_type="Brain"
 /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
 Pieter de Jong"
 BASE COUNT 90 a 63 c 52 g 143 t
 ORIGIN
 Query Match 0.6%; Score 23; DB 17; Length 348;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 TTGAAGAAATTTTCTCTTTT 594
 |||||
 Db 239 TTGAAGAAATTTTCTCTTTT 261

RESULT 24
 BQ488250 571 bp mRNA linear EST 07-JUN-2002
 LOCUS
 DEFINITION 39-E8847-006-003-M09-T3 Sugar beet MP1Z-ADIS-006 Lambda Zap II
 library Beta vulgaris cDNA clone M-9-3, mRNA sequence.
 ACCESSION BQ488250
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Beta vulgaris.
 Beta vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.

REFERENCE 1 (bases 1 to 571)
 AUTHORS Bellin,D., Werber,M., Theis,T., Weissshaar,B. and Schneider,K.
 TITLE EST sequencing, annotation and macroarray expression analysis of
 more than 3000 sugar beet cDNAs identifies genes with root-specific
 expression pattern
 JOURNAL Unpublished (2002)
 COMMENT ADIS DNA core facility at MP1Z
 Contact: Weissshaar B
 Max-Planck-Institute for Plant Breeding Research
 Carl-von-Linne Weg 10, 50829 Koeln, Germany
 Fax: 00492215062851
 Email: weissshaar@mpiz-koeln.mpg.de
 Seq primer: T3 /AATTACCCCTCACTAAAGG.
 High quality sequence stop: 571.
 Location/Qualifiers
 1. .571
 /organism="Beta vulgaris"
 /db_xref="taxon:161934"
 /clone="M-9-3"
 /clone_lib="Sugar beet MP1Z-ADIS-006 Lambda Zap II
 library"

/dev_stage="4 week old pot-grown plants"
 /note="Organ: shoot and root; Vector: pBluescript SK- from
 lambda Zap II; cDNA (lambda Zap-II) library from sugar
 beet, whole plant mRNA, prepared using the Stratagene
 UniZAP cDNA kit, cloning sites EcoRI-XhoI, primer sites
 and orientation:
 rev-T3-Saci-SK-EcoRI-GGCACGAGG-5pr-cDNA-polyA-XhoI-Kpnl-T7
 -uni."

BASE COUNT 172 a 103 c 128 g 168 t
 ORIGIN

Query Match 0.6%; Score 23; DB 14; Length 571;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2207 TTCTTTCAATGATGTTGCAAC 2229
 |||||
 Db 544 TTCTTTCAATGATGTTGCAAC 566

RESULT 25
 BB013867/c 265 bp mRNA linear EST 22-JUN-2000
 LOCUS
 DEFINITION BB013867 RIKEN full-length enriched, adult male testis (DH10B) Mus
 musculus cDNA clone 493047H22 3', mRNA sequence.

```

ACCESSION      BS013867
VERSION        GI:8137026
KEYWORDS       EST.
SOURCE         house mouse.
ORGANISM       Mus musculus

REFERENCE
AUTHORS        Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
               P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
               Hirozane,T., Hori,F., Ishi,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
               Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
               Kiyosawa,H., Kojima,Y., Kondo,S., Koye,S., Kurihara,C., Kusakabe,M.,
               Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
               Ono,T., Owa,C., Saito,E., Sakai,C., Sato,K., Shibata,K., Shibata,
               Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sobabe,Y., Sugahara,Y.,
               Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya
               T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
               Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
               M., Muramatsu,M. and Hayashizaki,Y.
TITLE          RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL        Unpublished (2000)
COMMENT        Contact: Yoshihide Hayashizaki
               Laboratory for Genome Exploration Research Group, RIKEN Genomic
               Sciences Center(GSC), Yokohama Institute
               The Institute of Physical and Chemical Research (RIKEN)
               1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
               Tel: 81-45-503-9222
               Fax: 81-45-503-9216
               URL: genome.gsc.riken.go.jp/
               Email: genome.gsc.riken.go.jp/
               Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
               N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
               Thermostabilization and thermoactivation of thermostable enzymes by
               trehalose and its application for the synthesis of full length
               cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
               Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
               Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
               Y. and Hayashizaki,Y.
               Automated filtration-based high-throughput plasmid preparation
               system. Genome Res. 9 (5), 463-470 (1999)
               Carninci,P. and Hayashizaki,Y.
               High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
               19-44 (1999)
               Please visit our web site (http://genome.rtc.riken.go.jp) for
               further details.

FEATURES       source
               Location/Qualifiers
               1..265
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="4930472H22"
               /clone_lib="RIKEN full-length enriched, adult male testis
               (DH108)"
               /sex="male"
               /tissue_type="testis"
               /dev_stage="adult"
               /lab_host="DH108"
               /note="Site.1: Sali; Site.2: BamHI; cDNA library was
               prepared and sequenced in Mouse Genome Encyclopedia
               Project of Genome Exploration Research Group in Riken
               Genomic Sciences Center and Genome Science Laboratory in
               RIKEN. Division of Experimental Animal Research in Riken
               contributed to prepare mouse tissues. 1st strand cDNA was
               primed with a primer [5'
               GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
               prepared by using trehalose thermo-activated reverse
               transcriptase and subsequently enriched for full-length by
               cap-trapper. Second strand cDNA was prepared with the
               primer adapter of sequence [5'
               GAGAGAGATTCGAGTTAATTAATTAATCCGCCCCCCCCC 3']. cDNA
               was cloned into the XhoI and BamHI sites. Vector: a
               modified pBluescript KS(+) after bulk excision from Lambda

```

```

BASE COUNT      91 a 37 c 39 g 98 t
ORIGIN          FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."

Query Match      0.6% Score 22; DB 10; Length 265;
Best Local Similarity 100.0%; Pred. No. 9.4; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 574 GAAGAAATTTTCTTTCTTTT 595
Db 150 GAAGAAATTTTCTTTCTTTT 129

RESULT 26
AV045785/c
LOCUS          AV045785 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
DEFINITION     1700048K22, mRNA sequence.
ACCESSION     AV045785
VERSION       AV045785.2 GI:4865450
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 308)
AUTHORS        Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
               Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
               A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
               Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H., Owa,C.,
               Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sobabe,Y., Sugahara
               Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y., Tomimaga,N.,
               Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
               Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE          RIKEN Mouse ESTs
JOURNAL        Unpublished (1999)
COMMENT        Contact: Chie Owa
               Genome Science Laboratory
               RIKEN
               3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
               Tel: 81-298-36-9345
               Fax: 81-298-36-9098
               Email: genome-res@rtc.riken.go.jp
               Thermostabilization and thermoactivation of thermostable enzymes by
               trehalose and its application for the synthesis of full length cDNA
               (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
               Transcriptional sequencing: A method for DNA sequencing using RNA
               polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
               Please visit our web site (http://genome.rtc.riken.go.jp) for
               further details.
               Location/Qualifiers
               1..308
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="1700048K22"
               /clone_lib="Mus musculus adult C57BL/6J testis"
               /sex="male"
               /tissue_type="testis"
               /dev_stage="adult"
               /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
               ) with a modified polylinker; Site.1: Not I; Site.2: Eco
               RI; 1st strand cDNA was primed with a Not I - oligo(dT)
               primer [5'
               TGTACCAATCTGAAGTGGAGCGCGCCGCAATGGTTTTTTTTTTTTTTTTTTTTTTT
               T 3']; double-stranded cDNA was ligated to Eco RI
               adaptors (Pharmacia), digested with Not I and cloned into
               the Not I and Eco RI sites of the modified pT73 vector.
               RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
               constructed and normalized by Bento Soares and M. Fatima
               Bonaldo."
BASE COUNT      107 a 44 c 52 g 105 t
ORIGIN          TGTACCAATCTGAAGTGGAGCGCGCCGCAATGGTTTTTTTTTTTTTTTTTTTTTTT

```


Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dt18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT73-pac vector. The library contained approximately 7.2 x 10⁵ recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 175 a 87 c 207 t

Query Match 0.6%; Score 22; DB 10; Length 580;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1049 CTTGGGCGAGCATTAAATGCT 1070

DB 489 CTTGGGCGAGCATTAAATGCT 468

RESULT 31
AV332932
LOCUS
DEFINITION
AV332932 RIKEN full-length enriched, adult male medulla oblongata
Mus musculus cDNA clone 6330543D05 3' similar to X64146
G domesticus mRNA BT61, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
AV332932.1 GI:6372984
house mouse.

REFERENCE
AUTHORS
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
Location/Qualifiers
1..275
/organism="Mus musculus"

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6330543D06"

/clone_lib="RIKEN full-length enriched, adult male medulla oblongata"

/sex="male"

/tissue_type="medulla oblongata"

/dev_stage="adult"

/lab_host="DH10B"

/note="Site 1: Sall; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTTCVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGATTAATTAATTAATTCCTCCCTCC sequence [5' GAGAGAGATTCGATTAATTAATTAATTCCTCCCTCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from lambda FLC 1. Cloning sites, 5' end: Sall; 3' end: BamHI"

BASE COUNT 60 a 43 c 46 g 126 t
ORIGIN

Query Match 0.6%; Score 21; DB 10; Length 275;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 576 AGAAATTTTCTTTCTTTTAT 596

DB 227 AGAAATTTTCTTTTAT 247

RESULT 32
AV207607/c

LOCUS
DEFINITION
AV207607 RIKEN full-length enriched, adult male testis muscle
CDNA clone 1700096A08 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
AV207607.1 GI:6148460
house mouse.
Mus musculus

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>

FEATURES
Location/Qualifiers
1..275
/organism="Mus musculus"

Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)

Itoh, M., Kitsumai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation System. *Genome Res.* 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

```
1. .282
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="1700096A08"
/clone="RIKEN full-length enriched, adult male testis"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGGCGCAATTAATTCAGTTAATTAATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. "
```

```
BASE COUNT      99 a  40 c  45 g  98 t
ORIGIN
```

```
Query Match      0.6%; Score 21; DB 9; Length 282;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 575 AAGAAATTTTTCCTTTTA 595
|||||
Db 161 AAGAAATTTTTCCTTTTA 141
```

RESULT 33

LOCUS

```
DEFINITION      BG379160 c
                  BG379160 284 bp mRNA linear EST 12-MAR-2001
                  UI-R-BT1-bnk-e-09-0-UI.s1 UI-R-BT1 Rattus norvegicus cDNA clone
                  BG379160
                  BG379160 284 bp mRNA linear EST 12-MAR-2001
                  BG379160 284 bp mRNA linear EST 12-MAR-2001
                  BG379160 284 bp mRNA linear EST 12-MAR-2001
```

```
ACCESSION      BG379160.1 GI:13303632
VERSION        BG379160.1
KEYWORDS       EST.
```

SOURCE

```
ORGANISM       Rattus norvegicus
                Norway rat.
                Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Rattus.
```

REFERENCE

AUTHORS

```
TITLE          1 (bases 1 to 284)
                Bonaldo, M.F., Lennon, G. and Soares, M.B.
                Normalization and subtraction: two approaches to facilitate gene
                discovery
                Genome Res. 6 (9), 791-806 (1996)
```

JOURNAL

MEDLINE

```
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
```

TITLE

```
JOURNAL
Unpublished (2000)
Contact: Yoshihide Hayashizaki
```

FEATURES

Source

```
1. .284
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BT1-bnk-e-09-0-UI"
/clone_lib="UI-R-BT1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The library
UI-R-BT1 is a subtracted library derived from a mixture of
the following tissues: hippocampus, thalamus, mid-brain,
medulla, corpus striatum, cerebral cortex and testis. For
a detailed description of the library from which this
clone was derived, please visit our web site at
rategen.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_SEQ=None found"
```

```
BASE COUNT      78 a  63 c  58 g  85 t
ORIGIN
```

```
Query Match      0.6%; Score 21; DB 12; Length 284;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 2770 AGCAATGCCATGTCGTATT 2790
|||||
Db 103 AGCAATGCCATGTCGTATT 83
```

RESULT 34

LOCUS

```
DEFINITION      BB446629 RIKEN full-length enriched, 9 days embryo Mus musculus
                  BB446629 323 bp mRNA linear EST 19-JUL-2000
                  CDNA clone D030056M06.3, similar to U39066 Murine MAP kinase kinase
                  6C mRNA, mRNA sequence.
                  BB446629
                  BB446629 323 bp mRNA linear EST 19-JUL-2000
                  BB446629 323 bp mRNA linear EST 19-JUL-2000
                  BB446629 323 bp mRNA linear EST 19-JUL-2000
```

```
ACCESSION      BB446629
VERSION        BB446629.1 GI:9289241
KEYWORDS       EST.
```

SOURCE

ORGANISM

```
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

REFERENCE

AUTHORS

```
1 (bases 1 to 323)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Oho, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suganara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
M., Muramatsu, M. and Hayashizaki, Y.
```

```
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
```

University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) Seq primer: M13 Forward POLYA=Yes.

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sueniro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuina, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

```
1..323
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D030056M06"
/clone_lib="RIKEN full-length enriched, 9 days embryo"
/dev_stage="9 days embryo"
/lab_host="DH10B"
/notes="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATTCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pluscript KS(+) after bulk excision from Lambda FLC I."
```

BASE COUNT

```
81 a 75 c 57 g 110 t
```

Query Match 0.6%; Score 21; DB 10; Length 323;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1293 TTTATTTTGTGTCGACAGG 1313

```
|||||
```

Db 56 TTTATTTTGTGTCGACAGG 76

RESULT 35

AV202914

LOCUS 381 bp mRNA linear EST 26-JUL-1999

DEFINITION AV202914 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA

clone yk556l2 5', mRNA sequence.

ACCESSION AV202914

VERSION AV202914.1 GI:5586685

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae

; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 381)

AUTHORS Kohara, Y., Shin, I. T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuki, H., Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A., Sano

M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and Nomoto, H.
Expressed genes in C.elegans
Unpublished (1999)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykoha@elab.nig.ac.jp.
Location/Qualifiers

FEATURES

source

```
1..381
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk556l2"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
```

BASE COUNT 118 a 66 c 77 g 119 t 1 others

ORIGIN

Query Match 0.6%; Score 21; DB 9; Length 381;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGCGAG 649

```
|||||
```

Db 33 TCATATGAAATATATGCGAG 53

RESULT 36

BE581464/c

LOCUS 426 bp mRNA linear EST 09-MAY-2001

DEFINITION KQ46b12.y1 TEN95TM-SSR Strongyloides stercoralis cDNA 5', mRNA

sequence.

ACCESSION BE581464.

VERSION BE581464.1 GI:9832406

KEYWORDS EST.

SOURCE

ORGANISM Strongyloides stercoralis.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Panagrolaimidae; Strongyloidea; Strongyloidea.

1 (bases 1 to 426)

AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,

Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,

Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., R.,

Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,

M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,

Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and

Wilson, R.

The Washington Univ. Nematode EST Project, 1999

Unpublished (1999)

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

The library was constructed by Dr. Thomas Nutman and colleagues of

NIH (tnutman@nih.gov). DNA sequencing by: Washington

University Genome Sequencing Center St. Louis.

High quality sequence stop: 411.

Location/Qualifiers

1..426

/organism="Strongyloides stercoralis"

/strain="Rhabditiform larvae obtained from gerbils"

/db_xref="taxon:6248"

/clone_lib="TEN95TM-SSR"

/lab_host="XL-1 Blue MRF" (Stratagene)"

/note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1: EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3 rhabdiform larvae which had been isolated from gerbils experimentally infected with larvae originally isolated from experimentally infected dogs. cDNA was constructed and using adaptors, was cloned unidirectionally into the vector from the EcoRI site to the XhoI site. The library has an unamplified titer of 1 x 10E5 pfu/ml and an amplified, undiluted titer of 9 x 10E11 pfu/ml. The average insert size of the unamplified library is 675 bp (range, 100-1700)."

BASE COUNT 168 a 47 c 76 g 135 t
ORIGIN

Query Match 0.6%; Score 21; DB 10; Length 426;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1509 ATAAAAATTGAATTCCTG 1529
|||||
DB 250 ATAAAAATTGAATTCCTG 270

RESULT 37
LOCUS A0498746 440 bp DNA linear GSS 28-APR-1999
DEFINITION HS_5075_B2_G09_SP6E RPCI-11 Human Male BAC Library Homo sapiens
ACCESSION A0498746
VERSION A0498746.1 GI:4698858
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 440)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380389
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 651 row: N column: 18
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 440.
Location/Qualifiers
1..440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=651 Col=18 Row=N"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 159 a 88 c 101 g 90 t 2 others

ORIGIN

Query Match 0.6%; Score 21; DB 17; Length 440;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2788 ATTTCAGACAGAGAACCTGA 2808
|||||
DB 386 ATTTCAGACAGAGAACCTGA 406

RESULT 38
BG705217/c 484 bp mRNA linear EST 07-MAY-2001
LOCUS 602688174F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4820322 5',
DEFINITION mRNA sequence.
ACCESSION BG705217 GI:13979333
VERSION BG705217.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 484)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshitsuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10725 row: d column: 19
High quality sequence stop: 398.
Location/Qualifiers
1..484
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4820322"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."

FEATURES
source

BASE COUNT 140 a 76 c 115 g 153 t
ORIGIN

Query Match 0.6%; Score 21; DB 12; Length 484;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 GAAGAAATTTTCTTTTCTTTT 594
|||||
DB 476 GAAGAAATTTTCTTTTCTTTT 456

RESULT 39
BI693834 500 bp mRNA linear EST 18-SEP-2001
LOCUS 603342457F1 NC1_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5370135 5',
DEFINITION mRNA sequence.

```

ACCESSION      BI693834
VERSION        BI693834.1  GI:15656450
KEYWORDS
SOURCE
ORGANISM       Mus musculus
               house mouse.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
               1 (bases 1 to 500)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
               Email: cgabs-remail.nih.gov
               Tissue Procurement: Gilbert Smith, Ph.D.
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM1942 row: a column: 16
               High quality sequence start: 5
               High quality sequence stop: 500.

FEATURES       source
               Location/Qualifiers
               1..500
               /organism="Mus musculus"
               /strain="FVB/N-3"
               /db_xref="taxon:10090"
               /clone="IMAGE:5370135"
               /clone_lib="NCI_CGAP_Mam2"
               /tissue_type="tumor, biopsy sample"
               /dev_stage="5 months"
               /lab_host="DR10B"
               /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
               Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
               Library constructed by Life Technologies. Investigator
               providing samples: Gilbert Smith, NIH"

BASE COUNT    165 a 101 c 103 g 131 t
ORIGIN
Query Match      0.6%; Score 21; DB 13; Length 500;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2694 TGGACCATGATGTCACAAATTG 2714
      |||||||||||||||||||
Db 237 TGGACCATGATGTCACAAATTG 217

RESULT 40
AZ930706
LOCUS         504 bp DNA linear GSS 01-APR-2001
DEFINITION   474.dhz57g01.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces
              unisporus genomic clone 474.dhz57g01.s1, DNA sequence.
ACCESSION    AZ930706
VERSION      AZ930706.1  GI:13501615
KEYWORDS     GSS.
SOURCE       Saccharomyces unisporus.
ORGANISM     Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
              Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE    Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
              W.R., Waterston,R.H. and Johnston,M.
              Surveying Saccharomyces genomes to identify functional elements by
              comparative DNA sequence analysis
              Unpublished (2001)
JOURNAL      Contact: Johnston M
              Department of Genetics
              Washington University Medical School
              Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
              Tel: 314 362 2735
              Fax: 314 362 7855

```

```

Email: mj@genetics.wustli.edu
Class: random plasmid subclone.
Location/Qualifiers
      1..504
      /organism="Saccharomyces unisporus"
      /strain="NRRL Y-1556 (CBS 398)"
      /db_xref="taxon:27294"
      /clone="474.dhz57g01.s1"
      /clone_lib="Saccharomyces unisporus NRRL Y-1556"
      /note="Random genomic sequence"

BASE COUNT    174 a 71 c 75 g 184 t
ORIGIN
Query Match      0.6%; Score 21; DB 17; Length 504;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2270 AGAATATTTATTCCTCAATGGA 2290
      |||||||||||||||||||
Db 172 AGAATATTTATTCCTCAATGGA 192

RESULT 41
BI038396/c
LOCUS         521 bp mRNA linear EST 14-JUN-2001
DEFINITION   RC5-NT0266-180101-021-A06 NT0266 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BI038396
VERSION      BI038396.1  GI:14445022
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 521)
REFERENCE    Dias Neto,E., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
              Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
              Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
              Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
              20202663
              Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-NT0266-
              180101-021-A06&t3=2001-01-18&t4=1)
              Seq primer: puc 18 forward
              High quality sequence stop: 483.

FEATURES       source
               Location/Qualifiers
               1..521
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="NT0266"
               /dev_stage="Adult"
               /note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI;
               Site_2: SmaI; A mini-library was made by cloning products
               derived from ORESTES PCR (O.S. Letters Patent application
               No. 196,716 - Ludwig Institute for Cancer Research)
               profiles into the pUC 18 vector. Reverse transcription of
               tissue mRNA and cDNA amplification were performed under
               low stringency conditions."

BASE COUNT    163 a 89 c 148 g 121 t
ORIGIN

```

Query Match 0.6%; Score 21; DB 13; Length 521;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTCTTTTCTTTTCTTA 599
 |||||
 Db 215 AATTTTCTTTTCTTTTCTTA 195

RESULT 42
 BI041199
 LOCUS
 DEFINITION RC5-NT0266-220201-013-E01 NT0266 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BI041199
 VERSION BI041199.1 GI:14447825
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 529)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL PROC.Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC5&t2=RC5-NT0266-220201-013-E01&t3=2001-02-22&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 528.
 Location/Qualifiers
 1..529
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0266"
 /dev_stage="Adult"
 /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 106 a 143 c 86 g 194 t
 ORIGIN

Query Match 0.6%; Score 21; DB 13; Length 529;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTCTTTTCTTTTCTTA 599
 |||||
 Db 306 AATTTTCTTTTCTTTTCTTA 326

RESULT 43
 AQ85038/c

LOCUS AQ85038 630 bp DNA linear GSS 30-JAN-2000
 DEFINITION RPCI-23-322G7.TJ RPCI-23 Mus musculus genomic clone RPCI-23-322G7, DNA sequence.
 ACCESSION AQ85038
 VERSION AQ85038.1 GI:6818243
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 630)
 AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levis,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 322 row: G column: 7
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..630
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-322G7"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 179 a 129 c 123 g 199 t
 ORIGIN

Query Match 0.6%; Score 21; DB 17; Length 630;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2915 AAAGAAATACCCAGACTTCATG 2935
 |||||
 Db 428 AAAGAAATACCCAGACTTCATG 408

RESULT 44
 AZ866201/c
 LOCUS AZ866201 637 bp DNA linear GSS 21-FEB-2001
 DEFINITION 2M0176B12R Mouse 10Kb plasmid UUCG1M library Mus musculus genomic clone UUCG2M0176B12 R, DNA sequence.
 ACCESSION AZ866201
 VERSION AZ866201.1 GI:13067272
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 637)

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0176 row: B column: 12
Seq primer: CACACAGGAAACGCTATGACC
Class: plasmid ends
High quality sequence stop: 637.

FEATURES Location/Qualifiers
source
1. .637
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="GUGC2M0176B12"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, P-"
/note="Vector: PWD42nv; Purified genomic DNA from Mus musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g1147321141g1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 205 a 140 c 108 g 184 t

ORIGIN

Query Match 0.6%; Score 21; DB 17; Length 637;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 619 ACTTCAGCTTCATATGAAA 639
|||||

Db 385 ACTTCAGCTTCATATGAAA 365
|||||

RESULT 45
BH098015 755 bp DNA linear GSS 19-JUL-2001

LOCUS RPCI-24-228N6.TJ RPCI-24 Mus musculus genomic clone RPCI-24-228N6, DNA sequence.

DEFINITION BH098015.1 GI:14919361

ACCESSION BH098015

VERSION BH098015.1

KEYWORDS house mouse, house musculus

SOURCE GSS.

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 755)

REFERENCE

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0176 row: B column: 12
Seq primer: CACACAGGAAACGCTATGACC
Class: plasmid ends
High quality sequence stop: 637.

AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M., Tsegaye, G., Geer, K., Kroll, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P. and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished (1999)

COMMENT Other GSSs: RPCI-24-228N6.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong. Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 228 row: N column: 6
Seq primer: SP6
Class: BAC ends.

FEATURES Location/Qualifiers
source
1. 755
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-228N6"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 175 a 181 c 137 g 262 t

ORIGIN

Query Match 0.6%; Score 21; DB 17; Length 755;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
|||||

Db 144 AATTTTTCCTTTTATCTA 164
|||||

RESULT 46
BH608326/c 765 bp DNA linear GSS 15-DEC-2001

LOCUS BH608326 BOGYJ73TF BOGY Brassica oleracea genomic clone BOGYJ73, DNA sequence.

DEFINITION BH608326.1 GI:17860772

ACCESSION BH608326

VERSION BH608326.1

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 765)

REFERENCE Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M. Whole genome shotgun sequencing of Brassica oleracea Unpublished (2001)

AUTHORS Other GSSs: BOGYJ73TR
Contact: Chris Town

TITLE TIGR

JOURNAL 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF

```

Class: sheared ends.
Location/Qualifiers
1. .765
/organism="Brassica oleracea"
/strain="TOL000DH3"
/db_xref="taxon:3712"
/clone_lib="BOGY"
/clone="BOGYJ73"
/notes="vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"
277 a 141 c 140 g 207 t

BASE COUNT
ORIGIN

Query Match 0.6%; Score 21; DB 17; Length 765;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1514 AAATTGAATTTCTTGCAATT 1534
|||||
Db 690 AAATTGAATTTCTTGCAATT 670

RESULT 47
BG985828/c
LOCUS
DEFINITION
5160t3 NICHZ zebrafish normalized I Danio rerio cDNA clone 5160,
mRNA sequence.
BG985828
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 805)
Kudo, T., Tsang, M., Fukriede, N.A., Chen, X., Dedekian, M., Clarke
, C.J., Kiang, A., Schultz, S., Epstein, J.A., Toyama, R. and Dawid, I.B.
A gene expression screen in zebrafish embryogenesis
Genome Res. 11 (12), 1979-1987 (2001)
21588053
Contact: David IB
Laboratory of Molecular Genetics
National Institute of Child Health and Human Development, NIH
Bldg. 6B, Room 413, 9000 Rockville Pike, Bethesda, MD 20892 USA
Tel: 301 496 4448
Fax: 301 496 0243
Email: idawidenh.gov
Seq primer: T3.
Location/Qualifiers
1. .805
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5160"
/clone_lib="NICHZ zebrafish normalized I"
/dev_stage="bud to 10 somite stage embryos"
/notes="Vector: pBluescript KS+; Site_1: NotI; Site_2: SalI
; RNA was reverse-transcribed to first strand cDNA using
SuperScriptII reverse-transcriptase and tagged oligo-dT
primer which contains several restriction sites including
a NotI site.
gactagtcttagatcgatcgatcgagcggccgccttttttttttttttttttt. Second
strand DNA was synthesized by E. coli DNA polymerase I in
combination with E. coli RNase H and E. coli DNA ligase.
Double stranded cDNA was ligated with SalI adapter. These
cDNAs were cloned into the SalI/NotI site of pBluescript
KS+ and transformed into E. coli Electromax DH10B by
electroporation."
292 a 120 c 103 g 290 t

BASE COUNT
ORIGIN

Query Match 0.6%; Score 21; DB 13; Length 805;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES
source
1. .765
/organism="Brassica oleracea"
/strain="TOL000DH3"
/db_xref="taxon:3712"
/clone_lib="BOGY"
/clone="BOGYJ73"
/notes="vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"
277 a 141 c 140 g 207 t

BASE COUNT
ORIGIN

Query Match 0.6%; Score 21; DB 17; Length 808;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2807 GATATGGCCATGAGTGATCCA 2827
|||||
Db 704 GATATGGCCATGAGTGATCCA 724

RESULT 49
BF621359
LOCUS
DEFINITION
HVSME0006K07f Hordeum vulgare seedling shoot EST library
HVCNDA0001 (Cold stress) Hordeum vulgare cDNA clone HVSME0006K07f,
mRNA sequence.
BF621359
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
1 (bases 1 to 954)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R., Choi, D.W.
, Fenton, R.D. and Main, D.

```

```

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1934 AAATATGAATCAGACACATA 1954
|||||
Db 196 AAATATGAATCAGACACATA 176

RESULT 48
BH558896
LOCUS
DEFINITION
BOHLC24TR BOHL Brassica oleracea genomic clone BOHLC24, DNA
sequence.
BH558896
VERSION
KEYWORDS
SOURCE
ORGANISM
Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 808)
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_SSs: BOHLC24TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1. .808
/organism="Brassica oleracea"
/strain="TOL000DH3"
/db_xref="taxon:3712"
/clone="BOHLC24"
/clone_lib="BOHL"
/notes="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"
238 a 194 c 187 g 199 t

BASE COUNT
ORIGIN

Query Match 0.6%; Score 21; DB 17; Length 808;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2807 GATATGGCCATGAGTGATCCA 2827
|||||
Db 704 GATATGGCCATGAGTGATCCA 724

RESULT 49
BF621359
LOCUS
DEFINITION
HVSME0006K07f Hordeum vulgare seedling shoot EST library
HVCNDA0001 (Cold stress) Hordeum vulgare cDNA clone HVSME0006K07f,
mRNA sequence.
BF621359
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
1 (bases 1 to 954)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R., Choi, D.W.
, Fenton, R.D. and Main, D.

```

TITLE
Development of a genetically and physically anchored EST resource
for barley genomics: Morex cold-stressed seedling shoot cDNA
library

JOURNAL
Unpublished (2001)

COMMENT
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
total hq bases = 201
Seq primer: AATTAACCCCTCAATAAGGG
High quality sequence start: 110
High quality sequence stop: 461.

FEATURES
source
1..954
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEa0006X07f"
HVCDA0001 (Cold stress)
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 50C for 2 days. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, and 600000 pfu
were in vivo excised to give pBluescript SK(-) cDNA
phagemids. These steps were performed in the TJ Close
laboratory at the University of California, Riverside
(Choi, Close, Fenton). Phagemids were plated and picked at
the Clemson University Genomics Institute (CUGI) (Begum,
Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations
, DNA sequencing and sequence analysis were performed at
CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main
). The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 166 a 248 c 213 g 325 t 2 others

ORIGIN
Query Match 0.6%; Score 21; DB 12; Length 954;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 AAATTTTCTTTCTTTATCT 598
|||||
Db 871 AAATTTTCTTTCTTTATCT 891

RESULT 50
CNS02GIX/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
137B20 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL196386
VERSION
AL196386.1 GI:7834536
KEYWORDS
GSS: genome survey sequence.
SOURCE
Tetraodon nigroviridis.

ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1009)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 1009)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish tetraodon nigroviridis
Unpublished
3 (bases 1 to 1009)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
Location/Qualifiers
1..1009
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="137B20"
/note="Genoscope sequence ID : COAG137DA10SP1-end :
PUC-ori"
PUC-ori
359 a 172 c 149 g 326 t 3 others

BASE COUNT 359 a 172 c 149 g 326 t 3 others

ORIGIN
Query Match 0.6%; Score 21; DB 17; Length 1009;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 582 TTTTTCCTTTTATCTAGTG 602
|||||
Db 238 TTTTTCCTTTTATCTAGTG 218

Search completed: November 6, 2002, 03:25:53
Job time : 4579 secs

FEATURES		Location/Qualifiers	
source	1..3731	/organism="unknown"	
BASE COUNT	1064 a	669 c	849 g 1149 t
ORIGIN			
Query Match 100.0%; Score 3731; DB 6; Length 3731; Best Local Similarity 100.0%; Pred. No. 0; Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	GAATATTTCTTACTTACTTACCAGGGAATGACTCATCTACCTCCCTCAAGTCTTTGTGT	60
Db	1	GAATATTTCTTACTTACTTACCAGGGAATGACTCATCTACCTCCCTCAAGTCTTTGTGT	60
Qy	61	GTGTGATAATAAATTTGGTTGTCTTCAGTTTCAGTCACACTGCTGGGTAGTTTTAT	120
Db	61	GTGTGATAATAAATTTGGTTGTCTTCAGTTTCAGTCACACTGCTGGGTAGTTTTAT	120
Qy	121	TTTGCATAAATTCAGGGGGTATTCAGTTGGTGTAGCAATTTGAAAGTCGAACTGCACCT	180
Db	121	TTTGCATAAATTCAGGGGGTATTCAGTTGGTGTAGCAATTTGAAAGTCGAACTGCACCT	180
Qy	181	GGAAATTTGGCTACATGGGAAGACAAATTCAGTTTTCGGATTCCTTATCTTCTCTGC	240
Db	181	GGAAATTTGGCTACATGGGAAGACAAATTCAGTTTTCGGATTCCTTATCTTCTCTGC	240
Qy	241	GGAAAGTGGTTAAGTCATCTTAGAGAAATATACAGGATATGGAACGTATGTGCATTGGA	300
Db	241	GGAAAGTGGTTAAGTCATCTTAGAGAAATATACAGGATATGGAACGTATGTGCATTGGA	300
Qy	301	GGTTAAACAGTCCAAAGAGGATCTAGAGCATTTGCCAAAGTTCAATTTGCCGACACAT	360
Db	301	GGTTAAACAGTCCAAAGAGGATCTAGAGCATTTGCCAAAGTTCAATTTGCCGACACAT	360
Qy	361	AAGTGCCTGACAAATCATCTTTGGCTATATACAGGCTGTATTTGGCTCTTCTTATTT	420
Db	361	AAGTGCCTGACAAATCATCTTTGGCTATATACAGGCTGTATTTGGCTCTTCTTATTT	420
Qy	421	GAAGGCTTTGGAAATGAAGAACTGATATGTGCCAACTCGGGGCATATGTGGATCAGATGGA	480
Db	421	GAAGGCTTTGGAAATGAAGAACTGATATGTGCCAACTCGGGGCATATGTGGATCAGATGGA	480
Qy	481	TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG	540
Db	481	TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG	540
Qy	541	TACAGAACTTTCAATTCGAATTTGGCATTTGAAGAAATTTTTTCTTTTATCTAG	600
Db	541	TACAGAACTTTCAATTCGAATTTGGCATTTGAAGAAATTTTTTCTTTTATCTAG	600
Qy	601	TGGTTACGCTGACTATTAACCTTCAGCTTTCATATGAAGAAATATATGGCAGTGTGCTCCA	660
Db	601	TGGTTACGCTGACTATTAACCTTCAGCTTTCATATGAAGAAATATATGGCAGTGTGCTCCA	660
Qy	661	TCGTCCATATGGTCAAAATGCTCAGTTTTCCTCCATACAGTTATTTGGTCTCTCTCGGAT	720
Db	661	TCGTCCATATGGTCAAAATGCTCAGTTTTCCTCCATACAGTTATTTGGTCTCTCTCGGAT	720
Qy	721	CTATAAGAGACTTGAAGAACTCCCTGTATAGCTTCTTAAAGGAACTCCCTGATCAGTG	780
Db	721	CTATAAGAGACTTGAAGAACTCCCTGTATAGCTTCTTAAAGGAACTCCCTGATCAGTG	780
Qy	781	GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCAAGTATATGTTTGA	840
Db	781	GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCAAGTATATGTTTGA	840
Qy	841	GTTCGGTAGGGGTTCGTCTTCCAAATTCGAGGAAAGTTTTTCCACTATCCGAAAG	900
Db	841	GTTCGGTAGGGGTTCGTCTTCCAAATTCGAGGAAAGTTTTTCCACTATCCGAAAG	900
Qy	901	TGAAGCAATATTAATTTACAGACTGGTTTCACTTTTTCGTCTCTCAAAATTCGGCTCT	960
Db	901	TGAAGCAATATTAATTTACAGACTGGTTTCACTTTTTCGTCTCTCAAAATTCGGCTCT	960

Qy	961	GGTTCCCAATGTCCAGCCTCCGGAAGAAATTTCAATTCCTACAGATTTTGTCAAAAT	1020
Db	961	GGTTCCCAATGTCCAGCCTCCGGAAGAAATTTCAATTCCTACAGATTTTGTCAAAAT	1020
Qy	1021	TAGTTCTTTGGTACACATGGATGCATACCTGGCCAGCATTAATGTCTACTTTTCG	1080
Db	1021	TAGTTCTTTGGTACACATGGATGCATACCTGGCCAGCATTAATGTCTACTTTTCG	1080
Qy	1081	ATTAGTTGATCCCTCGAAGAGAGAAATGTGCATGCATTTAGAGCAATGCTTACGAAACTGTA	1140
Db	1081	ATTAGTTGATCCCTCGAAGAGAGAAATGTGCATGCATTTAGAGCAATGCTTACGAAACTGTA	1140
Qy	1141	CTATATAAAGGAGTGTGTATGATCCCGTGAAGTGGCTCACTGAGCAGTATGATGGTA	1200
Db	1141	CTATATAAAGGAGTGTGTATGATCCCGTGAAGTGGCTCACTGAGCAGTATGATGGTA	1200
Qy	1201	TCTCAAGGGTACACAACCTCCCAAAATCTCGCTCCATCACTTTAGATGATGGGTGGTGA	1260
Db	1201	TCTCAAGGGTACACAACCTCCCAAAATCTCGCTCCATCACTTTAGATGATGGGTGGTGA	1260
Qy	1261	TGTAAGAAGGGTCTCTAGTAAACACCAATGCAAAAGTTTATTTTTTGGTCCAGAGTTAATGT	1320
Db	1261	TGTAAGAAGGGTCTCTAGTAAACACCAATGCAAAAGTTTATTTTTTGGTCCAGAGTTAATGT	1320
Qy	1321	TTCCAATCGGGTCTCTCGCAATTTATTCGAACATAGATAACTTTCTCGTGTCTTT	1380
Db	1321	TTCCAATCGGGTCTCTCGCAATTTATTCGAACATAGATAACTTTCTCGTGTCTTT	1380
Qy	1381	TGTTGATGAGGAGTGGGAGAACTGTATTTCTACAGACTTATACCAAAAGCAAGTACTGG	1440
Db	1381	TGTTGATGAGGAGTGGGAGAACTGTATTTCTACAGACTTATACCAAAAGCAAGTACTGG	1440
Qy	1441	AAGTGTGTGTCAGACAAACATCTATGAGAGATCTTATCAACTCTCGGAAAGGCTTGT	1500
Db	1441	AAGTGTGTGTCAGACAAACATCTATGAGAGATCTTATCAACTCTCGGAAAGGCTTGT	1500
Qy	1501	AATTTGTGATAAAAAATTTGAATTTCTTCATCTTCGACGCCAGTTGCGGGAATAATTC	1560
Db	1501	AATTTGTGATAAAAAATTTGAATTTCTTCATCTTCGACGCCAGTTGCGGGAATAATTC	1560
Qy	1561	AGTGTGATGTTTGGCATCAAGACCTGGCCCTTACTGCAAAATGATATAAGAGCTTGGATGG	1620
Db	1561	AGTGTGATGTTTGGCATCAAGACCTGGCCCTTACTGCAAAATGATATAAGAGCTTGGATGG	1620
Qy	1621	TGATTTTTCGCAGATCAAGAAATGTGCGAAATATGCTGCCAGACTTGGTCAATCTTTGG	1680
Db	1621	TGATTTTTCGCAGATCAAGAAATGTGCGAAATATGCTGCCAGACTTGGTCAATCTTTGG	1680
Qy	1681	TTCCCTCCAGAGAGCTTTGAGTGTCTTAGGCATGAGTTGAAGTTATTCGCGATGTA	1740
Db	1681	TTCCCTCCAGAGAGCTTTGAGTGTCTTAGGCATGAGTTGAAGTTATTCGCGATGTA	1740
Qy	1741	GGTTCATGGAACACGATATGCTTTTCTGATGGAATTTGTAATAATATCTGGTACTTTGC	1800
Db	1741	GGTTCATGGAACACGATATGCTTTTCTGATGGAATTTGTAATAATATCTGGTACTTTGC	1800
Qy	1801	TGATAGAGTTGGCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGATTCGTA	1860
Db	1801	TGATAGAGTTGGCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGATTCGTA	1860
Qy	1861	TGGTGGATATAAAGGTGTTTGGTGGTGTGATCCGGATTCAATGAAGTTGCTTCTTTGAG	1920
Db	1861	TGGTGGATATAAAGGTGTTTGGTGGTGTGATCCGGATTCAATGAAGTTGCTTCTTTGAG	1920
Qy	1921	AAAGAGCATGTCGAATATGAATCAGACACATAAAGTTAGATGCTTGGATGAGCAA	1980
Db	1921	AAAGAGCATGTCGAATATGAATCAGACACATAAAGTTAGATGCTTGGATGAGCAA	1980
Qy	1981	ATATCAGCCTTCTTAAATTCGTCACATGATTAAGCTTCTTGTCTACACTTGGAGTGA	2040
Db	1981	ATATCAGCCTTCTTAAATTCGTCACATGATTAAGCTTCTTGTCTACACTTGGAGTGA	2040

QY 2041 AGATGAAGTTCGACAGAGCAAGCAAGAGGAGCTGTAGATCAGCTTGATGCTATCTTGA 2100
Db 2041 AGATGAAGTTCGACAGAGCAAGCAAGAGGAGCTGTAGATCAGCTTGATGCTATCTTGA 2100
QY 2101 TGATTCCTTTGAAGGACAGAGGAGCTTTGGAATTTGATGCTCTCTCGAGAGAACACATATAT 2150
Db 2101 TGATTCCTTTGAAGGACAGAGGAGCTTTGGAATTTGATGCTCTCTCGAGAGAACACATATAT 2150
QY 2161 TCTCAAGCAATGCTAAACTGTTGATATAGGCTGATGCTGAGCCCTTCTTCAATGAT 2220
Db 2161 TCTCAAGCAATGCTAAACTGTTGATATAGGCTGATGCTGAGCCCTTCTTCAATGAT 2220
QY 2221 GTTGCAAACTTCCCGCATCCAGTTCCTGATTTGCGGACTAGATCAAGAATATTTAT 2280
Db 2221 GTTGCAAACTTCCCGCATCCAGTTCCTGATTTGCGGACTAGATCAAGAATATTTAT 2280
QY 2281 TCCAAATGGAAGCAATGATGGATGTTGGATCAATCCAGAACCTTGGATATGTTCA 2340
Db 2281 TCCAAATGGAAGCAATGATGGATGTTGGATCAATCCAGAACCTTGGATATGTTCA 2340
QY 2341 GGTGTTTCTCAGTTTACTGCTGAGATGGAGAGTTTCTGACGATTTTACATCCATT 2400
Db 2341 GGTGTTTCTCAGTTTACTGCTGAGATGGAGAGTTTCTGACGATTTTACATCCATT 2400
QY 2401 TAATCAACAGCAGATCCACACAGTAATTTCAATCTGAAGGAAATGTTGTTTGA 2450
Db 2401 TAATCAACAGCAGATCCACACAGTAATTTCAATCTGAAGGAAATGTTGTTTGA 2450
QY 2461 AAATCCATGCTGATCCTGCTGATATTCGTTTAAAGGCTGTAAATGTTCCAGCGCT 2520
Db 2461 AAATCCATGCTGATCCTGCTGATATTCGTTTAAAGGCTGTAAATGTTCCAGCGCT 2520
QY 2521 GCACACATGTTAGATGTTGTTGATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580
Db 2521 GCACACATGTTAGATGTTGTTGATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580
QY 2581 ATGTTCTGGAGTATTTGGATGGGATATCTACTTGTGTTGGATCAAGACATGAT 2640
Db 2581 ATGTTCTGGAGTATTTGGATGGGATATCTACTTGTGTTGGATCAAGACATGAT 2640
QY 2641 CCCGCAAGGCAAGTCCAGCGATGGATATCCCTCCAGCACCCAGCATACAGTTGGACCA 2700
Db 2641 CCCGCAAGGCAAGTCCAGCGATGGATATCCCTCCAGCACCCAGCATACAGTTGGACCA 2700
QY 2701 TGATGTCACATTTGAGGAGTTGAAGAGTACTTCCACACATATATTTGGAATGACAGTTT 2760
Db 2701 TGATGTCACATTTGAGGAGTTGAAGAGTACTTCCACACATATATTTGGAATGACAGTTT 2760
QY 2761 GGGATCATAGCAATGCCATGCTGATTTGCGAGAGAACTGATATGGCCATGAG 2820
Db 2761 GGGATCATAGCAATGCCATGCTGATTTGCGAGAGAACTGATATGGCCATGAG 2820
QY 2821 TGATCCATGCAAAACTTGTGAGCTCTTTCAATTTGCAAGTGCAGTTCGAAAGACTGG 2880
Db 2821 TGATCCATGCAAAACTTGTGAGCTCTTTCAATTTGCAAGTGCAGTTCGAAAGACTGG 2880
QY 2881 TGTTCCTGCTCAATACATCTGATGCGGCTTAAGATTAACAGATCCAGATTCATGGATAA 2940
Db 2881 TGTTCCTGCTCAATACATCTGATGCGGCTTAAGATTAACAGATTCATGGATAA 2940
QY 2941 GCCGCAAGACCTGATATCTCAGAAAGATTTATTGGAAGCTTTTTCAGGAAAGTGA 3000
Db 2941 GCCGCAAGACCTGATATCTCAGAAAGATTTATTGGAAGCTTTTTCAGGAAAGTGA 3000
QY 3001 GGACAAAGCAGCTCAGCTAGCTCTATGCGGACCTTCAAGAGATGTTTCAAGGATC 3060
Db 3001 GGACAAAGCAGCTCAGCTAGCTCTATGCGGACCTTCAAGAGATGTTTCAAGGATC 3060
QY 3061 ATATGATGCTGATATGGAAGTTGATGATTTGAAGATTTACATTTGACGAAGCTTTTGACTA 3120
Db 3061 ATATGATGCTGATATGGAAGTTGATGATTTGAAGATTTACATTTGACGAAGCTTTTGACTA 3120
QY 3121 CAAACTGAATATGACAAACAGCTGGGTAAATTTAATGGACTACTATGGCATAAAACAGA 3180

Db 3121 CAAACTGAATATGACAAACAGCTGGGTAAATTTAATGGACTACTATGGCATAAAACAGA 3180
QY 3181 GGCTGAATATCTTAGTGGTGGCATTATGAAGGCATCAAAAACCTTTTACCGCAGAAAAGA 3240
Db 3181 GGCTGAATATCTTAGTGGTGGCATTATGAAGGCATCAAAAACCTTTTACCGCAGAAAAGA 3240
QY 3241 TGCTGAGGCCATTAGTGTCTGCTGAGGGGCTTTGAGGAAGGAGGAGAGCCGTGTTCAA 3300
Db 3241 TGCTGAGGCCATTAGTGTCTGCTGAGGGGCTTTGAGGAAGGAGGAGAGCCGTGTTCAA 3300
QY 3301 GAGCGCTAATATAGATGATGACATGTTTACAAAAGGCTTCGGCTTGGTACCACGTTTACATA 3360
Db 3301 GAGCGCTAATATAGATGATGACATGTTTACAAAAGGCTTCGGCTTGGTACCACGTTTACATA 3360
QY 3361 TCATCTACATATTTGGGTTGCTACATCAGGCGTTGAAAAGAGCTCATTTTCAATAGCTT 3420
Db 3361 TCATCTACATATTTGGGTTGCTACATCAGGCGTTGAAAAGAGCTCATTTTCAATAGCTT 3420
QY 3421 TCCTGTGTGTTTATGACCGAGCTAATCCAGATTAAGAAGGACAAAGCAGTAAACAGGCC 3480
Db 3421 TCCTGTGTGTTTATGACCGAGCTAATCCAGATTAAGAAGGACAAAGCAGTAAACAGGCC 3480
QY 3481 AGTTCTCAACTTGTCTCTCAGGCTCAACTGAGTCACAGATTAGTTGTAATGAGA 3540
Db 3481 AGTTCTCAACTTGTCTCTCAGGCTCAACTGAGTCACAGATTAGTTGTAATGAGA 3540
QY 3541 TTCCAGTCGAGCGTTAAGCTGATATATATAATGTAATAGGCTGTGATCATAGAAGAAC 3600
Db 3541 TTCCAGTCGAGCGTTAAGCTGATATATATAATGTAATAGGCTGTGATCATAGAAGAAC 3600
QY 3601 TGTATGCAATTTGACTACCTTTTGTCTTAAACTGATGAGCTGCAACATATATGC 3660
Db 3601 TGTATGCAATTTGACTACCTTTTGTCTTAAACTGATGAGCTGCAACATATATGC 3660
QY 3661 AGTACTCTAAGAACAGATGATGAGCTAGTACTATATATGATGATTTGAGTTTCATC 3720
Db 3661 AGTACTCTAAGAACAGATGATGAGCTAGTACTATATATGATGATTTGAGTTTCATC 3720
QY 3721 TTTCTTCTAAA 3731
Db 3721 TTTCTTCTAAA 3731

RESULT 2

LERDRP LERDRP 3731 bp mRNA linear PLN 18-DEC-1998
LOCUS L.esculentum mRNA for RNA-directed RNA polymerase.
DEFINITION Y10403
ACCESSION Y10403.1 GI:4038591
VERSION RdRP gene; RNA-directed RNA polymerase.
KEYWORDS tomato.
SOURCE ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 3731)
AUTHORS Schiebel, W., Fellissier, T., Riedel, L., Thalmeir, S., Schiebel, R.,
Kempe, D., Lottspeich, F., Sanger, H. L. and Wassenecker, M.
TITLE Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from tomato
JOURNAL Plant Cell 10 (12), 2087-2101 (1998)
MEDLINE 99055198
PUBMED 9836747
REFERENCE 2 (bases 1 to 3731)
AUTHORS Wassenecker, M.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-1997) M. Wassenecker, Max-Planck-inst. fuer
Biochemie, Viridiforschung, Am Klopferspitze 18a,
Planegg-Martinsried, 82152, FRG
FEATURES Location/Qualifiers
1..3731

gene
CDS
/organism="Lycopersicon esculentum"
/strain="Rutgers"
/db_xref="taxon:4081"
194...3538
/gene="RdRP"
194...3538
/EC_number="2.7.7.48"
/codon_start=1
/product="RNA-directed RNA polymerase"
/protein_id="CAA71421.1"
/db_xref="GI:4038592"
/translation="MKTITLVNRRFLYSGSYLKAEMKTDIVQLRAYVDMQGITLNE
RAFAQVADINISADKIITLANNRFLYSGSYLKAEMKTDIVQLRAYVDMQGITLNE
CGTISDDKFLVAGSTEVSIQFGLKAKFFFFSSGADYKQLSYENIQVWVLRHYG
ONAFLLIQFGLAPRVYKRLNCSYFFKFTDDQWRTTDPSPVGLSSSLCLQFPR
RGVLRNPFESFHYAERENNIITLQGETFFVSOKSALVNPQPGIISIPYKILFKI
SSLVHGCIPGPAHVYFPLVDPBRNVACIEHLEKLYIKKCCYDVRVRLTEGYD
GIYKGRQPKSPSITDDGLVYVRRVLYVPCYVFCGPEVNSVNRNSEDINFL
RVSEVDENKLSITDLPKASTGSGVRINIERILSTLRKGFVIGDKRPELAFSSS
QLRDSNVFASRPLGTANDIRAMGDFGOIKNVAKYARLQSGFSRETLSVLRHE
IEVDPVYKHGTSYVSDGKISGDFAHKVSCKGLQYTPSAFQIRYGGYGVGVD
PDSMKLSLKSMSKYESDNIKLDVLGWSKIQPCYLNRLQITLLSLGVKDEVLEQKQ
KEAVDOLDALHDSLAQEALEMLSPGENTNLIKMLNGCYKPDAPFESMLQFRA
SKLLDLRTSRIRIPNGRTMMGCLDSRLLEYGOVFQGTGAGHGESDLPFRNRSR
STNSFLKGNVVAKNPCILHPGDIRVLKAVNVRALHHMVDVVPFQKRRHPNECS
GSDLDGIYFVCDQDKIPRQVQPMETPPAPSIQDHDVYIIEVEEYFTNYIVNDSL
GIILAHNVFADREPMAMSDPKLAEIIFSIADVDPFGVPAEIPSLRPEYEDFM
DKPDKTSYISERVIGLFRKVKDKPAQSSIAFTFRDVARSYDADMEYDGFEDYDE
AFDYKTEYDNKGLMDYVIGKTEAEILSGGIMKASKTFDRRDKDAISAVYRALRKE
ARAWFKRRNDIDMLPKASAWYHVYVHTPYWGCYNQGLKRAHFIISFPWCYVDQLIQIK
KDKARNRPNLNSLSRAQLSHRLVK"

BASE COUNT 1064 a 668 c 850 g 1149 t

ORIGIN

Query Match 100.0%; Score 3729.4; DB 8; Length 3731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3730; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAATATCTTTACTTACTTACCAGGATGAGTCACTACCTCCCTCAAGTCTTGTGT 60
DB 1 GAATATCTTTACTTACTTACCAGGATGAGTCACTACCTCCCTCAAGTCTTGTGT 60
QY 61 GTTGTGATAAATAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGGTAGTATTTAT 120
DB 61 GTTGTGATAAATAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGGTAGTATTTAT 120
QY 121 TTTCATTAACCTTCAGGGGTAATCCAGTTGGTGTAGCATTTGAACTGCACTGCACCT 180
DB 121 TTTCATTAACCTTCAGGGGTAATCCAGTTGGTGTAGCATTTGAACTGCACTGCACCT 180
QY 181 GGAATTTGGCTACATGGGAAAGACAATTCAGGTTTCGGATTCCTTATCTCTCTGC 240
DB 181 GGAATTTGGCTACATGGGAAAGACAATTCAGGTTTCGGATTCCTTATCTCTCTGC 240
QY 241 GGAAGTGGTTAAGTCATCTTTAGAGAATAATACAGGATATGGAACCTGATGTGCATTTGA 300
DB 241 GGAAGTGGTTAAGTCATCTTTAGAGAATAATACAGGATATGGAACCTGATGTGCATTTGA 300
QY 301 GGTAAACAGTCCAAAGAGGATCTAGAGCAATTTGCCAAAGTCAATTTGCCGACACAT 360
DB 301 GGTAAACAGTCCAAAGAGGATCTAGAGCAATTTGCCAAAGTCAATTTGCCGACACAT 360
QY 361 AAGTGTCTGACAAAATCATCACTTTGGCTAATAACAGGCTGATTTTGGCTCTCTTTATTT 420
DB 361 AAGTGTCTGACAAAATCATCACTTTGGCTAATAACAGGCTGATTTTGGCTCTCTTTATTT 420
QY 421 GAAGGCTTGGAAATGAAACTGATATTTGCCAACTCGGGGCATATGTGCATGAGTGA 480
DB 421 GAAGGCTTGGAAATGAAACTGATATTTGCCAACTCGGGGCATATGTGCATGAGTGA 480
QY 481 TGGCATTAACCTTTGATTTTCGATGTCAGATATCAGATGACAAGTTTCAGTGTGGGAAG 540
DB 481 TGGCATTAACCTTTGATTTTCGATGTCAGATATCAGATGACAAGTTTCAGTGTGGGAAG 540

DB 481 TGGCATAACTTTTGAATTTCCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG 540
QY 541 TACAGAAAGTTTCAATTCATTTGGCATTTGGATGGAAGAAATTTTTCCTTTTATCTAG 600
DB 541 TACAGAAAGTTTCAATTCATTTGGCATTTGGATGGAAGAAATTTTTCCTTTTATCTAG 600
QY 601 TGGTTCAGCTGACTATAAACTTCAGCTTTTCATATGAAAATATATGCGCAGGTTGTGCTCA 660
DB 601 TGGTTCAGCTGACTATAAACTTCAGCTTTTCATATGAAAATATATGCGCAGGTTGTGCTCA 660
QY 661 TGTTCATATGTTGTCAAAATGCTCAGTTTCTCTCATACAGTTTATTTGGTGTCTCTCGGAT 720
DB 661 TGTTCATATGTTGTCAAAATGCTCAGTTTCTCTCATACAGTTTATTTGGTGTCTCTCGGAT 720
QY 721 CTATAAGAGACACTTGAAGAACTCCTGTTTATAGCTTCTTTAAGGAAACCTCCTGATGATCAGTG 780
DB 721 CTATAAGAGACACTTGAAGAACTCCTGTTTATAGCTTCTTTAAGGAAACCTCCTGATGATCAGTG 780
QY 781 GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCTAGCTTATGTTTGA 840
DB 781 GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCTAGCTTATGTTTGA 840
QY 841 GTTCCGTAGGGGTGTCGCTTCCCAAATTTCCGAGGAAAGTCTTTTCCACTATGCGAAGG 900
DB 841 GTTCCGTAGGGGTGTCGCTTCCCAAATTTCCGAGGAAAGTCTTTTCCACTATGCGAAGG 900
QY 901 TGAACAACATATTAATTTACAGACTGGTTTTCACCTTTTTCGCTCTCTCAAAAATCGGCTCT 960
DB 901 TGAACAACATATTAATTTACAGACTGGTTTTCACCTTTTTCGCTCTCTCAAAAATCGGCTCT 960
QY 961 GGTTCCTCAATGTCACGCTCCCGAAGGAATTTCAATTCCTTCAAGATTTTGTTCAAAAT 1020
DB 961 GGTTCCTCAATGTCACGCTCCCGAAGGAATTTCAATTCCTTCAAGATTTTGTTCAAAAT 1020
QY 1021 TAGTTCTTTGGTACAGATGAGTGCATACCTGGCCAGCATTAATGTCTACTTTTTCG 1080
DB 1021 TAGTTCTTTGGTACAGATGAGTGCATACCTGGCCAGCATTAATGTCTACTTTTTCG 1080
QY 1081 ATTAGTTGATCTTCGAGGAGAAATGTCATGTCATGAGCATGCCCTTAGAGAAACTGTA 1140
DB 1081 ATTAGTTGATCTTCGAGGAGAAATGTCATGTCATGAGCATGCCCTTAGAGAAACTGTA 1140
QY 1141 CTATATAAGAGAGTGTCTGTTATGATCCCTGAGGTGGCTCAGTGCAGTATGATGGGTA 1200
DB 1141 CTATATAAGAGAGTGTCTGTTATGATCCGTGAGGTGGCTCAGTGCAGTATGATGGGTA 1200
QY 1201 TCTCAAGGTAGACACACTCCAAAATCTCCGTCCATCACTTTTAGATGATGGTGTGTA 1260
DB 1201 TCTCAAGGTAGACACACTCCAAAATCTCCGTCCATCACTTTTAGATGATGGTGTGTA 1260
QY 1261 TGTAAAGAGGGTCTTAGTAACACCATGCAAAAGTTTATTTTGTGTCAGAGGTTAATGT 1320
DB 1261 TGTAAAGAGGGTCTTAGTAACACCATGCAAAAGTTTATTTTGTGTCAGAGGTTAATGT 1320
QY 1321 TTCCAATCGGGTCTCCGCAATTTATTTCTGAACATAGATAACTTTCTTCGTGTTCTTT 1380
DB 1321 TTCCAATCGGGTCTCCGCAATTTATTTCTGAACATAGATAACTTTCTTCGTGTTCTTT 1380
QY 1381 TGTGTGATGAGGAGTGGGAGAACTGTATTTCTACAGACTTATTTACCAAAAAGCAAGTACTGG 1440
DB 1381 TGTGTGATGAGGAGTGGGAGAACTGTATTTCTACAGACTTATTTACCAAAAAGCAAGTACTGG 1440
QY 1441 AAGTGTGTGAGGACAACATCTATGAGGAGTCTTATCAACTCTCGCGAAGGCTTTGT 1500
DB 1441 AAGTGTGTGAGGACAACATCTATGAGGAGTCTTATCAACTCTCGCGAAGGCTTTGT 1500
QY 1501 AATTGTGTGATAAAAATTTGAAATTTCTTCATTCGAGCAGTTCGCGGATTAATTC 1560
DB 1501 AATTGTGTGATAAAAATTTGAAATTTCTTCATTCGAGCAGTTCGCGGATTAATTC 1560
QY 1561 AGTGTGGATTTTGTGATCAAGACCTGGCCTTTACTGCAATGATATAAGAGCTTGGATGGG 1620
DB 1561 AGTGTGGATTTTGTGATCAAGACCTGGCCTTTACTGCAATGATATAAGAGCTTGGATGGG 1620

QY 1621 TGAATTTTCGAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680
DB 1621 TGAATTTTCGAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680
QY 1681 TTCTCCAGAGAGACTTTGAGTGTCTTTAGGCATGAGATTGAAGTTATTCGCCGATGAAA 1740
DB 1681 TTCTCCAGAGAGACTTTGAGTGTCTTTAGGCATGAGATTGAAGTTATTCGCCGATGAAA 1740
QY 1741 GGTTCATGGAACCGAGCTATGCTTTTCGATGGAATGTTGTAATAATCTGGTGCATTTGC 1800
DB 1741 GGTTCATGGAACCGAGCTATGCTTTTCGATGGAATGTTGTAATAATCTGGTGCATTTGC 1800
QY 1801 TCATAGAGTTGCCCAAAATGTCGCTTCATATACCCATCTCTCTTCAGATTCGTTA 1860
DB 1801 TCATAGAGTTGCCCAAAATGTCGCTTCATATACCCATCTCTCTTCAGATTCGTTA 1860
QY 1861 TGGTGGATATAAAGGTGTGTGGTGTGTGATCCGGATTCATCAATGAAGTTGTCTTTGAG 1920
DB 1861 TGGTGGATATAAAGGTGTGTGGTGTGTGATCCGGATTCATCAATGAAGTTGTCTTTGAG 1920
QY 1921 AAAGAGCATGCGAATATGAATCAGACACATAAAGTTAGATGTCCTTGGATGGACAA 1980
DB 1921 AAAGAGCATGCGAATATGAATCAGACACATAAAGTTAGATGTCCTTGGATGGACAA 1980
QY 1981 ATATCAGCCTTGTATCTTAATCGTCAACTGATTACGCTCTTGTCTACACTTGGAGTAA 2040
DB 1981 ATATCAGCCTTGTATCTTAATCGTCAACTGATTACGCTCTTGTCTACACTTGGAGTAA 2040
QY 2041 AGATGAAGTTCTCAACAGAGCAAGCAAGAGCTGTAGATCAGCTTGCATCTATCTTGA 2100
DB 2041 AGATGAAGTTCTCAACAGAGCAAGCAAGAGCTGTAGATCAGCTTGCATCTATCTTGA 2100
QY 2101 TGATCTTTGAAGCACAGGAGGCTTTGGAAATGATGCTCTCTGGAGAGCACTAATAT 2160
DB 2101 TGATCTTTGAAGCACAGGAGGCTTTGGAAATGATGCTCTCTGGAGAGCACTAATAT 2160
QY 2161 TCTCAAGCAATGCTAACTGTGGTTATAGCCGTGATGCTGAGCCCTTTCTTCAATGAT 2220
DB 2161 TCTCAAGCAATGCTAACTGTGGTTATAGCCGTGATGCTGAGCCCTTTCTTCAATGAT 2220
QY 2221 GTTCAAAACCTTCGCGCATCCAAAGTTGCTCGATTTCGGACTAGATCAAGATATTTAT 2280
DB 2221 GTTCAAAACCTTCGCGCATCCAAAGTTGCTCGATTTCGGACTAGATCAAGATATTTAT 2280
QY 2281 TCCAAATGGAAGAACATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGTC 2340
DB 2281 TCCAAATGGAAGAACATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGTC 2340
QY 2341 GGTCTTTGCTCAGTTTACTGGTGTGCACATGGAGATTTCTGACGATTTACATCCATT 2400
DB 2341 GGTCTTTGCTCAGTTTACTGGTGTGCACATGGAGATTTCTGACGATTTACATCCATT 2400
QY 2401 TAATAACAGAGATCCCAACAGTAATTTCAATCTCAAGGAAATGTGGTTGTGCAAA 2460
DB 2401 TAATAACAGAGATCCCAACAGTAATTTCAATCTCAAGGAAATGTGGTTGTGCAAA 2460
QY 2461 AAATCCATGCTGCATCCTGGTGATATTCGTGTTTAAAGCGTGTAATGTTTCGAGCGCT 2520
DB 2461 AAATCCATGCTGCATCCTGGTGATATTCGTGTTTAAAGCGTGTAATGTTTCGAGCGCT 2520
QY 2521 GCACCATGGTATGATGTTGTATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580
DB 2521 GCACCATGGTATGATGTTGTATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580
QY 2581 ATGTCFGGAGTGATTTGGATGGGATATCTACTTGTGTTGGGATCAAGACATGAT 2640
DB 2581 ATGTCFGGAGTGATTTGGATGGGATATCTACTTGTGTTGGGATCAAGACATGAT 2640
QY 2641 CCGCCCAAGCAAGTCCAGCCGATGGAAATATCTCCAGCACCAGCATACAGTTGACCA 2700
DB 2641 CCGCCCAAGCAAGTCCAGCCGATGGAAATATCTCCAGCACCAGCATACAGTTGACCA 2700

QY 2701 TGATGTCAAAATGAGAAAGTTGAAGAGTACTTCCACCAACTATATTGTGAATGACAGTTT 2760
DB 2701 TGATGTCAAAATGAGAAAGTTGAAGAGTACTTCCACCAACTATATTGTGAATGACAGTTT 2760
QY 2761 GCGAATCATAGCAAAATGCCATGTCGTATTGTCAGACAGAACTGATATGCCATGAG 2820
DB 2761 GCGAATCATAGCAAAATGCCATGTCGTATTGTCAGACAGAACTGATATGCCATGAG 2820
QY 2821 TGATCCATGCAAAAAACTTGTCTGAGTCTCTTTCAATTTGCAGTGGACTTTTCCAAAGACTGG 2880
DB 2821 TGATCCATGCAAAAAACTTGTCTGAGTCTCTTTCAATTTGCAGTGGACTTTTCCAAAGACTGG 2880
QY 2881 TGTTCCTCGTGAATATACCATCTCAGTTCGCCCTTAAGAAATACCCAGACTTCATGATAA 2940
DB 2881 TGTTCCTCGTGAATATACCATCTCAGTTCGCCCTTAAGAAATACCCAGACTTCATGATAA 2940
QY 2941 GCCGGCAAGACCAAGTATATCTCAGAAAGAGTTATTGAAAGCTTTTCCAGGAAAGTAA 3000
DB 2941 GCCGGCAAGACCAAGTATATCTCAGAAAGAGTTATTGAAAGCTTTTCCAGGAAAGTAA 3000
QY 3001 GGCAAAAGCACCTCAGCTAGCTCTATCGGGACCTTCAAGAGATGTTGCAAGGAGATC 3060
DB 3001 GGCAAAAGCACCTCAGCTAGCTCTATCGGGACCTTCAAGAGATGTTGCAAGGAGATC 3060
QY 3061 ATATGATGCTGATATGGAAGTTGATGGATTGAAGATTACATTGACGAAAGCTTTTGACTA 3120
DB 3061 ATATGATGCTGATATGGAAGTTGATGGATTGAAGATTACATTGACGAAAGCTTTTGACTA 3120
QY 3121 CAAACTGAAATGACAAAGCTGGTAAATTAATGGACTACTATGGCATATAAAACAGA 3180
DB 3121 CAAACTGAAATGACAAAGCTGGTAAATTAATGGACTACTATGGCATATAAAACAGA 3180
QY 3181 GCGTGAATACTAGTGGTGGCATTAAGGCATCAAAACTTTTGACCGCAGAAAGA 3240
DB 3181 GCGTGAATACTAGTGGTGGCATTAAGGCATCAAAACTTTTGACCGCAGAAAGA 3240
QY 3241 TGCTGAGGCCATAGTGTGCTGTGAGGGCTTGAGGAAAGGAGGCAAGACCTTGGTTCAA 3300
DB 3241 TGCTGAGGCCATAGTGTGCTGTGAGGGCTTGAGGAAAGGAGGCAAGACCTTGGTTCAA 3300
QY 3301 GAGGCGTAAATGATAGATGACATGTTACCAAGGCTTCGGCTTGTACCAAGCTTACATA 3360
DB 3301 GAGGCGTAAATGATAGATGACATGTTACCAAGGCTTCGGCTTGTACCAAGCTTACATA 3360
QY 3361 TCATCTACATATTGGGGTGTCTACAAATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCTT 3420
DB 3361 TCATCTACATATTGGGGTGTCTACAAATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCTT 3420
QY 3421 TCCTCGGTGTTTATGACCAAGCTATTCAGATTAAAGAGGACAAAGACCTTAACAGGCC 3480
DB 3421 TCCTCGGTGTTTATGACCAAGCTATTCAGATTAAAGAGGACAAAGACCTTAACAGGCC 3480
QY 3481 AGTTCTCAACTTGTCTCTCAGGGCTCAACTGAGTCAAGATTAGTCTTGAATGAGA 3540
DB 3481 AGTTCTCAACTTGTCTCTCAGGGCTCAACTGAGTCAAGATTAGTCTTGAATGAGA 3540
QY 3541 TTCCAGTCGAGCGTTAAAGCTGATATATATTAATGATAGGGTGTGATCATAAGAAAC 3600
DB 3541 TTCCAGTCGAGCGTTAAAGCTGATATATATTAATGATAGGGTGTGATCATAAGAAAC 3600
QY 3601 TGTATGCACTTGTGACTACCTTTTCTTTAAACTGCAAGCTCAACATATATGC 3660
DB 3601 TGTATGCACTTGTGACTACCTTTTCTTTAAACTGCAAGCTCAACATATATGC 3660
QY 3661 AGTACTCTAAGAAACAGATGTACAGTAACTAATATGATGATTTGATTTGATTTTCATC 3720
DB 3661 AGTACTCTAAGAAACAGATGTACAGTAACTAATATGATGATTTGATTTGATTTTCATC 3720
QY 3721 TTTCTCTTAAA 3731
DB 3721 TTTCTCTTAAA 3731

```

RESULT 3
NTA011576      3505 bp      mRNA      linear      PLN 07-JAN-1999
LOCUS          Nicotiana tabacum mRNA for RNA-directed RNA polymerase.
DEFINITION
ACCESSION     AJ011576
VERSION       AJ011576.1 GI:4138281
KEYWORDS      drp gene; RNA-directed RNA polymerase.
SOURCE        common tobacco.
ORGANISM      Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; eustersids I; Solanales; Solanaceae; Nicotiana.
Wassenegger, M.
Direct Submission
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
Martinsried, FRG
2 (bases 1 to 3505)
Schiebel, W., Pelissier, T., Riedel, L., Thalmair, S., Schiebel, R.,
Kempe, D., Lottspeich, F., Saenger, H. L. and Wassenegger, M.
Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from Tomato
Unpublished
Location/Qualifiers
1..3505
/organism="Nicotiana tabacum"
/cultivar="Petit Havana SRI"
/db_xref="taxon:4097"
/tissue_type="leaves"
20..3370
/gene="RdRp"
20..3370
/gene="RdRp"
/codon_start=1
/evidence=experimental
/product="RNA-directed RNA polymerase"
/protein_id="CAA09697.1"
/db_xref="GI:4138282"
/translation="MGKTIQVGFYLLSAEAVKFLNHTNGVIVALEYVQSKGGR
RAFAKQFANKSAEFLIDLSKGLYIGSYLKWEMKTDIVOPRTYVEHMDGVTLNF
GCQIDSKKFAVLGSDVSIKFGILKKIYFFLSHASADYKLQLSYENTWQVYLHRPYG
QNAQFLLOLFGAPRIYKRLEDSCSYFFKETPDQDWRITDFTPSWIGLSSLOLEFR
NVLQPNFSEFSEFYKESMNQFILOTGFTFSQKALVPIVHLPEGLPKYILFKFI
SLIOLHGCPGLALNENFQLVDPNRNACIEHLEKLYLKECCYDPVRLTEQYD
EYLKGRQLPKSPITLDDGLVYRVVVTVPCKYFCGPPVNVSNVLRNYSIDINFL
RVSFYDEEKEKHSIDLLPRSTNGRTDIYERILSLNRNGFIIDKRFPEAFSSS
QLRDNVWFAFRPGLTANDIRTMGDFRQIRNVAKYARLGQSFSGSRETLSVGHE
VEVIPDVACSLHGTNYIFSDGIGKISADFAHRTIKCGLQYTPSSFOIRYGGYKQVA
VDPYSMSLSLRSKLSKYSNNIKLDVLGSKYQPCYNLRQLVLLSLGLVDDVLEQ
KONEAYDQDLAILHDSLKAOEALMLSPGENTNILEMNLNCGYMPDAPFELSMLOTF
RASKLLDLRTRIFIPNGRTWMGCLDESRTLEYGVFVQFSGAGROFFESHPND
SGSANDCFILKGNVAVKNCPLHPGDIRVLRADYDPAALHHMYDCVVFQKGRHPNE
CSGSDLDGDIYFCWDPDLIPRQVQSDYTPAPTQLDHDVITIEVEEYFTNYLND
SUGITIANHVFADREPDMASDPCKQLAFLSIADVDPKTPGPAEIPSLRPKEYPD
FMEKDPKTYPEKVIKGLFKRVKNKTPQASSIATFTDRVARKSYDSMDIVDGFEDYI
DEAFYKSYDNKGLMDYGYIKTEAELISGLGINKASKTFRDRKDAIGAVVACLR
KEARAFKRSRSDIDMLAKASAWYHVYHTYWGLYNEGLKRDHFIISFPWCYVDLIQ
IKKAKARKRPVPHLSLSLSQSRKLVW"
BASE COUNT    1020 a 622 c 785 g 1078 t
ORIGIN

Query Match      76.5%; Score 2852.8; DB 8; Length 3505;
Best Local Similarity 89.0%; Pred. No. 0;
Matches 3137; Conservative 0; Mismatches 362; Indels 27; Gaps 4;

QY 175 GCACCTTGGAAATTTGGCTACATGGAAGACAATTCAGGTTTCGGATTCCCTATCTTCT 234
Db 1 GCACCTTGGAAATTTGGCTACATGGAAGACAATTCAGGTTTCGGATTCCCTATCTTCT 60

QY 235 CTCTCGCGAAGTGGTAAAGTCATTCTTAGAGAAATATACAGGATATGGAACTGTATGTGC 294
Db 61 GTCTGCGAGACAGTCAGAGAAATCTTAGAGAAATATACAGGAAATGGAACTGTCTATGC 120

```



```
Db 3361 GAT-----ATGTAGCATTAAAGGAGCTATATA-ATGTAAAGGGTGTGA 3406
QY 3589 TCATAAGAAACTGTATGCAATGTTGACTACCTTTTCTCTTTTAAACTGCATGAAGCTG 3648
Db 3407 TCATAAGAAACTGTATGCAATGTTGACTACCTTTTCTCTTTTAAACT-----GGTG 3459
QY 3649 CAACATATACAGTACCTTAAGAAACAGATGACAGCTACAGTACT 3694
Db 3460 ATATATTTATGACACCTTTTAAACAGATGTTATAGCTAAGTAGT 3505

RESULT 4
PHY011979
LOCUS Petunia hybrida RdRp gene, partial. 4579 bp DNA linear PLN 07-JAN-1999
DEFINITION AJ011979
ACCESSION AJ011979
VERSION 1
KEYWORDS RdRp gene; RNA-directed RNA polymerase.
SOURCE Petunia x hybrida
ORGANISM Petunia x hybrida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
1 (bases 1 to 4579)
Wassenegger,M.
Direct Submission
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
Martinsried, FRG
2 (bases 1 to 4579)
Schleib,W., Pelissier,T., Riedel,L., Thalmair,S., Schiebel,R.,
Kempe,D., Lottspeich,F., Saenger,H.B. and Wassenegger,M.
Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from Tomato
Unpublished
Location/Qualifiers
1..4579
/organism="Petunia x hybrida"
/db_xref="taxon:4102"
/tissue_type="leaves"
1..4579
/gene="RdRp"
Join(<1..1451,2420..2570,3856..>4579)
/gene="RdRp"
/codon_start=1
/evidence-experimental
/product="RNA-directed RNA polymerase"
/protein_id="CAA09896.1"
/db_xref="GI:4138343"
/translaton="CIEHALEKLYNLEKCCYDPRWLTKQYDEYLRKROLPKSPSTIL
DGLMYVRRVLTPCKVYFLGPEVNVNRLVNYSEDIDNFLRVSVDFEWEKIFISMD
LLPRANRGTIDYIRLLSTLONGLVIGAKKEFLAFSSQLRDNVWVPASRGLT
AYDIRWMDGFOQIRNVAKYARLGOSFSSREILSYGRHEIKYIPDVEVHTNIVFS
DGKISADFAHRVARKCGLQHPSPAFQIRYGGYKGVAVDPYSPMKLSRLSKMLKYE
SNTKLDJVGWSKYQPCYLNRLQVLTMLSTLGVKDEYFKRQNEAVQLDAILHDSLKA
QEAALDSFENTNILEKMLNCGYLPDAEPFLSMQLQTPFASKLLDLRTRIFIPKG
RSMWGLDIESRTLEYGVVFOFTGAGROFYEESHFNDSGSANYNFIKGNVYVAKN
PCLPHGDYRVLKXAVDPALHHMVDVVFQPKGRPHNCSGSDLDLDIYFVCDPDL
IPROVQPDMDYTPASIQLDHDTVEEVEEYFTNMYNDGLGVIANHVVPADREPML
AMSDPIELAQLESIAVDFPKTGPVPEISQLRPKEIPDFNEKPDITTHSEKVIKGL
FRVKDKTQASSITFTRDVARDYDLEVDGFEDYIDEAFDYKSEYDKLGLNMD
YGIKTEAETLSGIMKASKTFDRRDAEIGAIVAVRSRKEARTWFKRSDIDDLIAK
ASAWYVAVHPTWYGCYNEGLKRDHFISFPWCYVD"
<1..1451
/gene="RdRp"
/number=1
1452..2419
/gene="RdRp"
/number=1
2420..2570
/gene="RdRp"
/number=2
2571..3855
```


CDS	8842.. 9714 /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F10B6.4" /protein_id="AAF79218.1" /db_xref="GI:8778209" /translation="MDSRYDDFIKETDHSILGLISMDVLOKKLYAEESORRURE QYLSRLRLEKEKYIDLVSEASMAQSLKFEVENQKIGSERDLVNOCKWKEK FLHQDSRLMEGNETDREAERAESEVEERKMSDEIKSR:ESSEDCLVDSL ASVSKDESLSLRIELEANSEDKCETLLRKWDQKPTQKEFVSLVSMVKRIEKEK ECUIMNLAKAEQVELVSEQNRELDRENKFLRQCSAERSHSGSNKFKRSIKMSSP IEKRIELSSQEFID" 11067..11906 /note="unknown protein; similar to EST gb AI995472.1" /codon_start=1 /evidence=not_experimental /product="F10B6.5" /protein_id="AAF79219.1" /db_xref="GI:8778210" /translation="MDDGFRNMGYYEPAATFKNLGLQLMSTIDRNTKPLGRDP NLMGPNQSYHGOEPPHMSYNINQOKDFNMLPVTATPNYGNVLEPETSASMQ MNLHHLOTEENVKLEELVQTKRK"NAKAGSTPKAKPRKPKDENSNANNNT NTRVPAKSDVLYINGVMDISGLPFICTCTGAPQOCIRWGCOWSACTNNIS MHPLPSTKRGARISGRKMSQGAFFKVLKLUASDGFNFGNPIDLKSHWARHGNTKV TIR" complement(13762..14781) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F10B6.6" /protein_id="AAF79250.1" /db_xref="GI:8778241" /translation="MKLWRAAVLVKDGPSLIAADDILTAAYVKATSHDELSIDTESA QFYIRVLSPSLKPLVLSISRVKRTSWAVALKMLMHGFFLCKTVAESIGRL PFLDSFGENSRIMSKSGFNLFVAYPAFLDRRSILPHDGNRHYNESESVILFLV IIRKMOIIVDSILRIPIGENMMIPVINEAMENVSEIWEIYCNICRRIRAEVLNVHS KIGKTADALKIVAKSMKGGLKFIKFECKDLGVSNAGEIPIFNVRPEADVIHLDE LVATMSESESAERTEIAEEEEEEETIKLSDLITLDHNEEAPASPVRVVVDIP DLISF" 16027..16533 /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F10B6.7" /protein_id="AAF79220.1" /db_xref="GI:8778211" /translation="MSTCVYRCMRNHAALGSAIDGCREYSOPSTGDLCAVGGCH RSYHRIDVSSPOINHTRPPTSLRRVKQLARLWKTAERNEEEDDTEETEEK MTVORRRKSKFTAQREAMKDYAAKLGWLKDKRALREIRVCEBGIGVTRHFRTW NNKKFYH" complement(17635..17815,17901..18019,18141..18194, 18300..18406,18547..18611,18742..18950)) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F10B6.8" /protein_id="AAF79249.1" /db_xref="GI:8778240" /translation="MEFKAYFIYWFQPRLOCEPVIYVFLCFQMCLKKKRIELFW YPBELHQPLTPKMYFIWVFLIRNMFEDKQVCKSSSTKGDVSVSGAELGFNE IVRPYVKSIGSFSGKSLGPKGSKVIFGOALKVGTSLTFVGEAVRDKAGNMIQKSK EQSLIVFSSESFDENVMNMKQSELCLVILAKIFGSIAYIAVYGVYDVARVLLPFV WKKDLGDSOSSONGSSDSDTDSF" complement(19071..20409,20499..20573,20648..20721, 20800..20887,20957..21108,21188..21356,21448..21906, 22006..22275,22367..22570,22840..22917)) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F10B6.9" /protein_id="AAF79248.1" /db_xref="GI:8778239"	8842.. 9714 /translation="MLEIESPTSLCFRTNTTCNALLRELQKIWVDIGSDAEKORMLM ELEKCELEIYRRKQEAQSLVSEAEIASLAAALGFVNSHPNKAKEGSK SLKEKLAARVPMLEDLRQKDERMFOVDIKAQIEKMSGEISGYSDOLNKTWVGSAL DQDQLTLRKINEVQTHLSLOKEDKDLNVLVDYVNEVHTLGVDFGQGVSEVHP SLHRTDHEQSTNISDDTLGDLHMIHLKTSREFOKVDKHACNAIILFYDVHEDA IAFSLVQLKDVAGSLFELNIMDTSEERTKFAASYVYVRSSESDITEENILSSET IEQVSAEVDCFNKLKASRMKELVMARTELENLCRLAHIEADTSTLSKSTALIDSGT KRLVPSSELLINIEHLINKIEEASRKEIIDRIDWLSACEEENWLEBYNODETRYS AGRGHVNLKHAERARITVNIKPSMVDNLIRKTLMEDETRKSFYLDGVRVLSLEDY KLTRQOEERKRYRQDKMODLIRRESIYSGKSPSRNSRKYNGVNGDASVPP TPRRNSAGATNNDIMTTPRSYSSHQNGYFKEVRRLSTAPLNFVAIPKEDSVSTYSV CGSEPDSPLYN" 26948,27454..27610,27713..27814,27911..28038, 28342..28470,28565..28632,28747..28969) /note="similar to acid phosphatase type 5 emb CAB63938.1; similar to EST gb N96048.1" /codon_start=1 /evidence=not_experimental /product="F10B6.10" /protein_id="AAF79221.1" /db_xref="GI:8778212" /translation="MTYIYRDTKITTKSTIPFLIFFLCFNSLNATLKHKPVNLVYF VTNLIIFSSHSSTAEELRLQLPSKTDGVSLVIGDMGRKGSINQSOVALQMEIGE VTNLIIFSSHSSTAEELRLQLPSKTDGVSLVIGDMGRKGSINQSOVALQMEIGE	Query Match 29.3%; Score 1094.4; DB 8; Length 132699; Best Local Similarity 60.7%; Pred. No. 1.5e-249; Matches 2100; Conservative 0; Mismatches 1181; Indels 177; Gaps 12;	QY 179 TTGGAAATTTGGTACATGGGAAGACAATTCAGGTTTCGGGATTCCTTCTCTCT 238 DB 65813 TTGATATAGAGAAAGATGGGAAGACAATTCAGGTTTCGGGATTCCTTCTCTCT 65754 QY 239 GCGGAAGTGGTAAAGTCATCTTAGAGAAATATACAGGATATGGAACCTGATGTGCATG 298 DB 65753 GCAGAGAAAGTTTGAAGATTTCTTGAAGACTCAGCTGTTGAGGACATCTTACCAATC 65694 QY 299 GAGGTTAAACAGTC---CAAAGAGGATCAGAGCATTTGCCAAAGATCAATTTCCCGAC 355 DB 65693 AAAGTTAGAACCAAGAAAGTGGTCCAGAGTCTATGCCATTTGTTCAATTCACATCT 65634 QY 356 AACATAAGTGTGTACAAAATCATCTTTGGCTAATAACAGGCTGTATTTGGCTCTTCT 415 DB 65633 GAGAGACATACGAGGCTGATCATCTGCGGACAGAGACGCTTTACTATGGAAGATCT 65574 QY 416 TATTGGAAGCTTGGGAATGAAACTGATATATGTCCAACTGC---GGGCATATGTGGAT 472 DB 65573 TATCTCAAGGCTTTTGAAGTTGAACAAGACATTTGTCCAAAGCCAGAGCTTCATTAC 65514 QY 473 CAGATGGATGGCATACTTTGAATTTCCGATGTAGATATCAGATGACAAAGTTTGCAGTG 532 DB 65513 ACCATCTCTGGTTTAAAGATGTTCTTTGGATGCAAGTCTCGACTAAGAAAGTTTAACT 65454 QY 533 TTGGGAAGTAC---AGAAAGTTTCAATCAATTTGGCATTTGGATTTGAAGAAATTTTTTC 589 DB 65453 CTATGGAGTGTCTCAAGATGTTTGTCTGTTTGAATTTGGATTTGGATGAGAAATTCGATTT 65394 QY 590 TTTTATCTAGTGGTTCAGCTCAGCTATAAATTCAGCTTTCATATGAAAATATATGGCAG 649 DB 65393 TCGTTTTCTTGATCAAAAGGATTTCTCTTGAGCTTTCGTATGAGAAATATATGGCAG 65334 QY 650 GTTGTCTCCATCGTCCATATGTCCTCAAAATGCTGATTTCTCTCTCATACAGTATTTGGT 709 DB 65333 ATTGATTTGCATTTCTCCGCAAGGAGCATCATCGAAAGTTTCTTTGTTGATTCAGGTGATTTGT 65274 QY 710 GTCCTCGCATATAGAGACTTTGAAACTCTCTTTTAGTCTTTT-----757 DB 65273 GTCCTCAAAATATTGAGAAAGAGGATCAACCTATCAATCTTTTGTGGAAATATAGGAT 65214 QY 758 -----AAGGAAACTCCTGATGATCAGTGGGTGAGGACAAACAGATTTCCCTCCA---TCT 808 DB 65213 TTTTACAGTGTAGGGTCTGATGAGCAATGATGAGAACTACAGATTTTACATCTTCTCTCT 65154 QY 809 TGGATAGGCTATCTCTAGCTTATGTTGACAGTTCCGAGGGGTGTCGCTTCCCAAT 868
-----	--	---	--	---

Db 65153 TGTATTGTCATCAACAGCGTTTGGTTAGAGCTTCCTGTTTCATCTTAATGTTCCCTGAT 65094
QY 869 TTCGAGGAAGTTTTCACACTATGCAGAAAGTGAACAATATATCTTTACACACTGGT 928
Db 65093 TTCAGAGAAATTTGCCAACTATGCAGAGCAGAGCGAGCTCTTTCTCATTTGAATCT 65034
QY 929 TTCACCTTTTTCGTCTCAAAAATCGGCTCTGTTTCCCAATGCCAGCTCCGGAAGGA 988
Db 65033 GGTTCAGCTATTCACTAATGCTAACACACTTGTCTCTGTTGATCCCTCCCGAGG 64974
QY 989 ATTTCAAATCCCTACAGATTTTGTTCAAAATAGTCTTTGGTACAGCATGGATGCA 1048
Db 64973 TTTAGTTTGGCTTTTGAGATCTTGTAAACCTTAACATATGGTGCAGAACGGTGTCTC 64914
QY 1049 CTTGGCCAGCATTAATATGTTCTTTTCCGATTAGTTGATCTCGAAGGAGAAATGTG 1108
Db 64913 TCCGGCCAGCTCTGTGCTTGGACTTTTATCGGTTGCTTAATCAGAAAAATATGATGA 64854
QY 1109 GCATGCATTGAGCTGCTTAGAAGAACTGTACTATATAAGGAGTGGCTGTATGATCCC 1168
Db 64853 GCTCTCATAGATCACTGTCTTGAAGAACTCTTTCACCTTGGAGAAATGTTGCTATGAGCCT 64794
QY 1169 GTGAGTGGCTCACTGAGCAGTATGAGGGTATCTCAAGGTTAGCAACCTCCAAAATCT 1228
Db 64793 GCTCATTTGGCTTCGATGAGTACAGAGTGGATATCGAAGGAAAGCTTCCACTGTCT 64734
QY 1229 CCGTCCATCACTTTAGATGATGGTGGTGTATGATAGAGGGTCTTAGTAACACCATGC 1288
Db 64733 CCTACAATATCTTGTATGATGAGTCTGTTATATATGATGATCGGTTCAAGTTACACGCA 64674
QY 1289 AAAGTTATTTTGTGTCAGAGAGTTAAATGTTTCCAAATCGGTTCCCGCAATATCT 1348
Db 64673 AGAGTCTATTTCTGACCGGAGTGAATGTTTCTAATCGTGGCTTCGACATTTACTCC 64614
QY 1349 GAAGACATAGATACTTCTTCGTGTTCTTTTGTGATGAGGAGTGGGAGAACTGTAT 1408
Db 64613 AAGTATATCAACAATCTCCAGGCTCTCATTTGTCATGAAGATCTTGAAGGTTCTG 64554
QY 1409 TCTACAGACTTATACCAAGCAAGTACTTGGAGTGGTCTCAGGCAACACATCTATGAG 1468
Db 64553 TCCATGATCTATCTCCAGCTCTTCCTACTCAAG-----AAGAACCAAGTTATACGAC 64500
QY 1469 AGGATCTTATCAACTCTCGGAAAGGCTTGTAAATGGTGATAAAAATTTGAATTTCT 1528
Db 64499 AGGATATATCTGTTCTCCGGATGGATGTCTAGAGCGATAGAAGTTTGGATTCTC 64440
QY 1529 GCATTTTCAGGAGCCAGTTTGGGGATTAATTCAGTGTGGAGTTTGCATCAAGACCTGGC 1588
Db 64439 GCCTTTTCTCCAGCCAGTTTGGGGAGACTCTGCTTGGATGTTTGACCAATAGATCGG 64380
QY 1589 CTTACTGCAATATGATAAGAGCTTGGATGGTGATTTTTCGAGATCAAGATGTCGCA 1648
Db 64379 ATTACAGCAGCAGATATCAGAGCTTGAAGGGTGATTTTGACCATATAGAATGTAGCT 64320
QY 1649 AAATATGCTGCCAGACTTGTGCAATCTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTT 1708
Db 64319 AAATATGCTGCAAGCTTGTGCAATCTTTAGCTCTCAAGAGAGACACTTAATGTTAGG 64260
QY 1709 AGGATGAGATGAGTATATCCGATGTAAGGT-----TCATGGAACCACTATGTC 1762
Db 64259 AGTATGAGATGAGTATGATCTGATGTTGAGATTATATCTTTAGGGACACGCTATGG 64200
QY 1763 TTTTCTGATGGAATTTGTAATAATCTTGTGACTTTGCTCATAGAGTTGGCTCAAAATGT 1822
Db 64199 TTTTCCGATGGAATAGGAAGATATCAGCTGAATTTGCTAGAAGGTAGCGAGGAATGT 64140
QY 1823 GGCCTT---CAATATACCCCATCTGCTTCCAGATTCGTTATCGTGTATATAAGGTGTT 1879
Db 64139 GGTCTTACAGAGTTTCTCCATCTGCTTTTCAAAATCCGTTTATGGCGTTTATAAGGAGTG 64080
QY 1880 GTGGGTGTTGATCCGGATTCATCAATCAAGTCTGCTTTTGAAGAGAGCATGTGCAATAT 1939

Db 64079 GTGGCTTTGATCCAAAACCTCATCAAGAACTGTCTCTGAGGAAGATATGACCAATTC 64020
QY 1940 GAATCAGACAAATAAAGTTAGATGTCCTTGGATGAGGAGCAAAATATCAGCCTTGTATCTT 1999
Db 64019 GAATCGGAGAACCAACAGCTGTGATGTTCTGGCGTGAGCAAGTACCAACCTTGTATATG 63960
QY 2000 AATCGTCAACTGATTACGCTCTTGTCTACACTTGGAGTGAAGATGAAGTTCTCGCAACAG 2059
Db 63959 AACAGACACTGATTACACTTTTGTCTACTCTTGGAGTTACGGACAGTGTGTTGAGAAG 63900
QY 2060 AAGCAAAAGAAAGCTGTAGATCAGCTTGTGCTATCTTGCATGATTTTGAAGGCACAG 2119
Db 63899 AACAAGAGGAAAGTTGGATCGTCTCGAGCAATCTTGCATCATCTTTTGAAGCTCAC 63840
QY 2120 GAGGCTTTGAAATGTGCTCTCTCGAGAGAACACTTAATTTCTCAAGCAATGCTAAAC 2179
Db 63839 GAGGCTCTTGGTTAATGCTCCAGGGAAACAAATATCTCAAGGCAATGATCTTIG 63780
QY 2180 TGTGGTTATAAGCCTGATGCTGAGCCCTTTCTTTCAATGATGTTGCAAACTTCCCGGCA 2239
Db 63779 TGTGGCTATAAAGCTGACGCTGAACCTTTCTTTCAATGATGCTTCAGAAATTCAGGCA 63720
QY 2240 TCCAAGTTGCTGATTTGGGACTAGATCAAGAAATATTTATCCAAATGGAAGCAATG 2299
Db 63719 TCTAAGTTGTTGAATCTACGCAAAAACAAAGATTTTCATTTCTGGTGAAGATCTATG 63660
QY 2300 ATGGGATGTTGGATGAATCCAGAACCTTTGGAATATGTTGTCAGTGTGTTTTCAGTTTACT 2359
Db 63659 ATGGGATGCTAGACAGACCCAGAACACTGGAGTATGTTGTCAGTGTAGTGCAGTATTC 63600
QY 2360 GGTGC--TGGACATGAGAGTTCCTGACAGTTTACATCCATTAATACACAGATCCA 2417
Db 63599 GATCCCATGAGCGCGGAAGGGGATTCATCATCAGCGGACCTGTTGTTGCCAAAAAC 63540
QY 2418 CCAACAGTAATTTCAATCTGAAGGAAATGTTGTTGCCAAAAATCCATGCTTGCATC 2477
Db 63539 CCATGCTGCATCCTGTCGCTGTCGCTGTTCTCAAGCTGTCAATGCCAGCTTTAAT 63480
QY 2478 CTGGTGATATTCGTGTTTAAAGGCTGTAATGTTGCGAGCGTGCACACATGTT----- 2532
Db 63479 CATATGGTGGACTGCTGTTGTTCCGCAAAAGGGTTGAGGTAAACATTTTCTATGCT 63420
QY 2533 -----AGATTGTTGTTGTTTCCCTCAGAAA----- 2557
Db 63419 TGATTCGTTGTTTTCGAGCTTTTATGTTTAACTCTCGAACITTTTCATTTGGAATCTT 63360
QY 2558 -----GGAAAAGACCTCATCCGAATGAATGTTCTGGAGTGATTTGGATGGGA 2607
Db 63359 GTGTTCTTTTGTGCATAGGCCACACCCAAATGAATGTTCTGGAGTGATTTAGATGGGA 63300
QY 2608 TATCTACTTTGTTGCTGGGATCAAGACATGATCCGCAAGGCAAGTCCAGCCGATGA 2667
Db 63299 TATTTACTTTGATGTTGGATCAAGAAATGTTCCGCAAGAACGTCGTAACCAATGA 63240
QY 2668 ATATCCTCAGCACCACCAATFACAGTTGGACCATGATGTCACAAATG----- 2714
Db 63239 CTACACTCTGAACCAACTCAAAATCTTGGATCATGATGTCACAAATGAGTTATAATAA 63180
QY 2715 ----- 2714
Db 63179 ATTTGATCACTTTGGTTATATTTTTTTTTTTTTTTGGCGTTAGGAGATGATTTGGCTTACCTT 63120
QY 2715 -----AGGAAGTTCAAGAGTACTTACCAACTATTTGTGAATCAAGATTTGGGAATC 2767
Db 63119 GGTTTCCAGGAAGTTGAAGAGTACTTTGCGAACTACATTTGTAATGATGATTTTAGGATC 63060
QY 2768 ATAGCAATGCCATGTCGTATTTTCAGACAGAACCTGATATAGCCATGATGATCA 2827
Db 63059 ATCGCAATGCTCATACCGCTTTTCGGATAAGAACCACTCAAGCGCTTTTGTAGTACCCA 63000
QY 2828 TGCAAAAATGCTGAGCTCTTTTCAATTTGCAAGTGGACTTTCCAAAAGACTGGTGTCTCC 2887
Db 62999 TGCATTTGAGCTTGCAAGAGATTTTCACTGCGGTAGATTTTCCCAAAACTGGTGTGCA 62940

```
QY 2888 GCTGAATACCATCTCAGTTCGCCCTAAAGAAATACCAGACTTCATGGATAAGCCGAC 2947
Db 62939 GCTGTATACCGACACATCTTACGTGAAGAGTATCCTGATTCATGGAGAACCGGAT 62880
QY 2948 AAGACCAGCTATATCTCAGAAAGAGTATTGGAAGCTTTTTCAGGAAGTGAAGGACAAA 3007
Db 62879 AAGCCACATATAGTCTGAAGAGCTGATTGGTAAGCTTTTTCAGAGAGGTAAAGAGCGA 62820
QY 3008 GCACCTCAGGCTAGCTCTATCGGCACCTTCACAAGAGATGTTGCAAGAGACATATGAT 3067
Db 62819 GCTCCACCGTTGATCTGATCAATCAATCACTGATGTTGGCTCAAGTCTTATGAT 62760
QY 3068 GCTGATATGGAAGTGTATGATTTGAAGATTAATCATTCAGCAAGCTTTTGACTACAAAAC 3127
Db 62759 AAAGACATGGAAGTGTACGGAATTTGAGAGTATGTTGATGAAGCTTTTCTACCAAGGCG 62700
QY 3128 GAATATGACACACAGCTGGGTAAATTAATGACTACTATGTCATGATGCAATAAAGAGGCTGAA 3187
Db 62699 AATTATGATTTCAAGTTAGTAAATTTGATGATTAATGATGATGATGATGATGATGATGAT 62640
QY 3188 ATACTTAGTGTGCATATTAAGAGGCTCAAAACATTTTACCGCGACAAAGATGCTGAG 3247
Db 62639 ATTCTTAGTGTGTATTAATGAGTGTCAAAAGTCAATTCACCAAGAGAGAGATGCAGAA 62580
QY 3248 GCCATTAGTGTCTGTGAGGCTTCAGCAAGGAGGCGACAGCCGTGTTCAAGAGCGGT 3307
Db 62579 TCGATTGAGAGCGGTTTCGGGCTTGAGGAAAGAAATTTGCTGTTTCAATGCTTCT 62520
QY 3308 AATGATATGATGACATGTTACCAAGGCTTCGGCTTGSTGATGACCGTTACATATCATCCT 3367
Db 62519 GAAGAAGAAATATGATGACCAAGGCTTCGGCTTGATCATGATTAAGTACCACCTCA 62460
QY 3368 ACATATTGGGTTGCTACATCATCAGGGTTCGAAGAGCTCATTTTCATGATTTCCCTGG 3427
Db 62459 AGTTACTGGGACTTTATATGAGGTTTGAATCGTGACCATTTCTTAAGTTTTCGGTGG 62400
QY 3428 TGTGTTTATGACCACTAAATCCAGATTGAAGAGGACAA 3465
Db 62399 TGCCTTTATGATAGCTTGTGAGGATCAAGAAGACCAA 62362
```

```
RESULT 5
ATH011977
LOCUS      1538 bp      DNA      linear      PLN 07-JAN-1999
DEFINITION Arabidopsis thaliana RDRP gene, partial.
ACCESSION  AJ011977
VERSION    AJ011977.1 GI:4127462
KEYWORDS   RDRP gene; RNA-directed RNA polymerase.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE  1 (bases 1 to 1538)
            Wassenegger, M.
            Direct submission
            Submitted (16-Oct-1998) Wassenegger M., Max-Planck-Institut of
            Biochemistry Viroldresearch, Am Klopferspitz 18A, 82152
            Martinsried, FRG
REFERENCE  2 (bases 1 to 1538)
            Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R.,
            Kempe, D., Lottspeich, F., Saenger, H.L. and Wassenegger, M.
            Isolation of an RNA-directed RNA polymerase-specific cDNA clone
            from Tomato
            Unpublished
FEATURES             Location/Qualifiers
     source           1..1538
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /tissue_type="leaves"
     gene             1..1538
```

```

/ gene="RDRP"
<1..1538
/ gene="RDRP"
/codon_start=1
/evidence-experimental
/product="RNA-directed RNA polymerase"
/protein_id="CAA09894.1"
/db_xref="GI:4375829"
/translation="IRAWGDFDHIRNVAARLGGSSSSRRLTNVRSDEIEVDP
VEISLGRVYVFDGIGKISAEFAKVKARKGLTFSPSAFOIRYGVYGVVAVDPNS
SKKLSLRKSMKFSFSEWTKLDYLAWSKYQPCYMMRQLITLLSTLOYDSVEKKQREV
VDRDLAITLHELEAHEALGMAPGENTNLRALILCGYKPDPEFSLSMQNFRAKSL
LELTKTRIFISGGRSMGCLDETLEIYGVVYQYSDMPRPFRRFITGVVYAKNP
CLHSGDVRVLAQVNPALNHWVDFVFPKGRPHNECSGSLDGDITFYFCWQELV
PPRSEPMDYTPETQILDHDTVEEVEEFANYIVNSLGTIANAHAFADKPPKLA
FSDPCIELAKKFSTAVDPKTVAAVPOHLYVKEYPDMPEKPKPTVYESKNVIGKLF
REKVERAPPLISIKSFTLDVASKYDKDMVEVDGFEYVDEAFYQKANDYDFKLGNDY
VGIKT"
BASE COUNT      442 a      279 c      381 g      436 t
ORIGIN
Query Match      20.5%; Score 765; DB 8; Length 1538;
Best Local Similarity 69.7%; Pred. No. 5,2e-171;
Matches 1104; Conservative 0; Mismatches 425; Indels 54; Gaps 3;
QY 1604 ATAAGAGCTTGGATGGGTGATTTTCGCGAGATCAAGATGTCGCAAAATATGCTGCCAGA 1663
Db 1 ATACGGCGCTGGATGGGTGATTTTGAACCATATAAGAAATGTAGCTAAATATGCTCAAGG 60
QY 1664 CTTGTGCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTTAGGCATGAGATTGAA 1723
Db 61 CTTGTCAATCTTTTAGCTCTTCAAGAGAGACACTTAATGTTAGGAGTGTAGAGATTGAA 120
QY 1724 GTATATCCCGATGTAAGGT-----TCATGGAACAGCATGTCTTTTCATGATGGAAT 1777
Db 121 GTGATCTCTGATGTTGAGATTATATCTTTAGGGACACCGTATGTCTTTCCGATGGAATA 180
QY 1778 GGTAAATATCTGGTGACTTTGCTCATAGATTGCTCAAAATGTGGGCTT---CAATAT 1834
Db 181 GGAAAGATATCAGCTGAATTTGCTAGAAAGGTAGCGAGGAATGTTCTTACAGAGTT 240
QY 1835 ACCCATCTGCTTCCAGATTCGTTATGTTGATATAAAGGTGTTGGTGGTGTGATCGG 1894
Db 241 TCTCCATCTGCTTTTCAAATCCGTTATGCGGTTTAAAGGAGTGGTGGCTGTGATCCA 300
QY 1895 GATTTCATCAATGAAGTGTCTTTGAGAAAGAGCATGTGAAATATGAATCAACACACATA 1954
Db 301 AACTCATCAAGAAACTGCTCTGAGGAGAGATGATGACCAATTCGAATCGAGAGACACC 360
QY 1955 AAGTTAGATGCTCTGGATGGAGCAAAATATCAGCCTTGTATCTTAATCGTCAACTGATT 2014
Db 361 AAGCTTGATGTTCTGCGCTGGAGCAAGTACCAACCTTGTATATGAACAGACAACTGATT 420
QY 2015 ACGCTCTTGTCTACACTTGGAGTGAAGATGAAGTCTCGAACAGACAAAGAAAGAGCT 2074
Db 421 ACACTTTGTCTACTCTTGGAGTTACGACAGCTGTTTGGAGAGAAACAAGGGAAGTT 480
QY 2075 GTAGATACGTTGATGCTATCTTCATGATTTCTTTGAGGACAGAGAGGCTTTGGAATG 2134
Db 481 GTGGATCGCTTCGACGCAATCTTGACTCATCTTTTGAAGGCTCAGCAGGCTCTTGGTTTA 540
QY 2135 ATGTCTCTGGAGAGAACACTAATATCTCAAGCAATGCTAAACTGTGTTATAAGCT 2194
Db 541 ATGGCTCAGGGGAAACACAAATATCTCAAGGCAATGATCTTGTGGGCTATTAAGCT 600
QY 2195 GATGCTAGCCCTTTCTTTCAATGATGTTGCAAACTTCCCGCGCATCCAAAGTTGCTGAT 2254
Db 601 GAGCGTGAACCTTCTCTTCAATGATGCTTCAGAAATTTTCAGGCGCATCTAAGTTGTGAA 660
QY 2255 TTGCGGACTAGATCAAGAATATTTATTCAAATGGAAGAACAAATGATCGGATGTTGGAT 2314
Db 661 CTACGGACAAAACAAAGGATTTTCATTTCTGTTGGAAGATCTATGATGGGATGCTAGAC 720
```


Db	92117	GAAGATCCATATTAATCAACTATATTTTAGGAGATCAATACAGATGATCAATGACCAAGCAACG	92117
Qy	794	GATTTCCCTCCATCTT---GGATAGGGGTATCTTCTAGCTTATGTTTGCAGTTCCGTAGG	850
Db	92177	GATTTTACTTCACTCTCTAGCATTTGGCAATCATATATTTTATGCTTTGAGGTACCACT	92236
Qy	851	GGGTGCTCTCTCCAAATTTTCGAGAAAGTTTTTTCACATGACGAGACGTGAAAACAAT	910
Db	92237	CGATGTGATCTCCCAACCAATTCGTGACTACTCTTTTACTATCATGAGTCAAAATCATGAC	92296
Qy	911	ATTACTTTACAGACTGGTTTTACCTTTTTTCGTCTCTCAAAAATGGCTCTCGTTCCCAAT	970
Db	92297	TTGAGTCCGGAGTGG---GGGTATCCATATTCAAAGTGATACTCGCTTTGTTCCGATT	92353
Qy	971	GTCCAGCCTCCGGAAAGAAATTTCAATTCCTACAGATTTTGTTCAAATTAGTCTCTTTG	1030
Db	92354	GTGAATCTCTGTGG-----TTATGTCCCTTACAGATACACTCTTCAAGATCAACCATTTG	92407
Qy	1031	GTACAGCATGATGATACACTCTGGGCCAGCATTAATATGCTACTTTTTCCGATTTAGTTGAT	1090
Db	92408	GTTCAGATGGGACACTTAGCGGCCGACACTTGATGACACCTTCTTCGCTCTGGTCAGC	92467
Qy	1091	CCTCGAAGGAAATGTGGCATGCAATGAGCATGCTTTAGAGAACTGTACTATATAAAG	1150
Db	92468	CCAGCTTTGTACTATTAGCATATAAAGCGTGCACCTTGAATATGATGTCATATTTGAAA	92527
Qy	1151	GAGTGCTGTTATGATCCGTCAGTGGCTCACTGAGCAGTATGATGGGTACTCTCAAGGGT	1210
Db	92528	AAGACCTGCTTGAATCCAAAGAATGGCTATCTGAACAANTACTCAAAATCCGGAGATCA	92587
Qy	1211	AGACAACCTCCAAAATCTCCGTCCATCACTTTAGATGATGGGTTGGTGATGTAAAGAAG	1270
Db	92588	CGCTACGTCCAACCATCACTAACATATCTCTGGAATGATGGATTGGTTATG-----	92639
Qy	1271	GTCCCTAGTAACACCATGCAAGTTTATTTTGTGGTCCAGAGGTTAATGTTTCCAATCGG	1330
Db	92640	-----TCTACAGTGATTTTTATGGACCTGAGATTAATGTTCTCCAATCGT	92684
Qy	1331	GTCTCCGCAATTTATCTGAAGACATAGATAAATCTTTCGTGTTCTTTTGTGTTGATGAG	1390
Db	92685	GTGTGCGGAATTTTCTCTCGACATAGAGAAATTCCTTCGCAATTCATTGTTGACGAG	92744
Qy	1391	GAGTGGGAGAACTGTATCTACAGACTTATACAAAGCAAGTACTGGAAGTGTGTG--	1448
Db	92745	GACTGCGAAGAGCTCCGTGCAACTGATTTATCCCCCGTCTGCTTCAGGACATGATGCA	92804
Qy	1449	-TCAGGACAAACATCTATGAGAGGATCTTATCAACTCTGCGAAAGGCTTTGTAATGGT	1507
Db	92805	AACAGACTGCTCTGTATAGAGAGTCTGTACGTCCTTTCAGATGCACTATTTGGT	92864
Qy	1508	GATAAAATTTGAATTTCTTGCAATTTTCATCGAGCCAGTTGCGGGATAATTCAGTGTGG	1567
Db	92865	GGCAAGACTTTGAGTTCTTCTGCTTTTCTTCAAGCCAGCTCCGAGATAACTCTGCAATG	92924
Qy	1568	ATGTTTGATCAAGACCTGGCTTACTGCAATGATATAAGAGCTTGGATGGGTGATTTT	1627
Db	92925	ATGTTTGCTCTCGCCAGGATTAGCTGCTAGCACATTTAGGACATGGATGGGGGATTTT	92984
Qy	1628	TCGAGATCAAGAAATGTCGAAAAATATGCTGCCAGACTTGGTCAATCTTTTGGTTCTCTC	1687
Db	92985	CGTAACATCAAAATGTAGCAAAATATGCTGCAAGACTTGGTCAGTCTTTTAGTTCTCTCA	93044
Qy	1688	AGAGAGACTTTGAGTGTCTTAGGCATGAGATTTGAAGTTATTCGCCGATGTAAAGGTTTAT	1747
Db	93045	ACCGAAACATTAAGGTTTCAGAATGATGAGGTGGAAGAAATCTCAGATATAAA---AAAT	93101
Qy	1748	GGAACCAAGCTATGCTCTTTTCTGATGGAAATGGTAAAAATATCTCGTGACTTTGCTCATAGA	1807
Db	93102	GGTACACAGCATGTTCTCTGATGGGATGTGAAGATATCATCTCGCTTTTGTCTAATGAA	93161
Qy	1808	GTTGCCCTAAATGTGGCCTTCAA---TATACCCCATCTGCTTTTCCAGATTCGTTATGGT	1864
Db	93162	TTGCCCATGAGTGCACACTGAAACGGTTTGCTCTCTGCTTTTCAATTAAGGTATGGT	93221

QY	1865	GGATATAAAGGTGGTGGGTGGTTCGCGGATTCATCAATCAAGTGTCTTTTGAAAG	1924
DB	93222	GTTTATAAAGGTGGTGGGTGGTTCGCGGATTCATCAATCAAGTGTCTTTTGAAAG	93281
QY	1925	AGCATGTCGAATATGAATCAGACAACATAAAGTTAGATGTCCTGGGATGAGCAAAAT	1984
DB	93282	AGCATGTCGAATATGAATCAGACAACATAAAGTTAGATGTCCTGGGATGAGCAAAAT	93341
QY	1985	CAGCCTTGTTATCTTAAATGGTCAACTGATACGCTCTTGTCTACACTTGGAGTGAAGAT	2044
DB	93342	CAGCAGGCTTCTTAAATGGCAGCTAAATCTCTTCTTTGACACTTGGAGTTAGAT	93401
QY	2045	GAAGTTCGACAGACAAGCAAGCACTGTAGATCAGCTTGATGCTCATCTTGCATCAT	2104
DB	93402	AGTGCTTTGAACAGACAGACAAGCAAGCACTGATCATAGTTGAACAAATGGTTACTGAT	93461
QY	2105	TCTTTGAAGCACAGGAGGCTTTGGAATTGATGTCCTCTGGGAGAGAACACTAATATCTC	2164
DB	93462	CCTCAAGCTGTCTATCGAAGCAATTGAACCTTATGCCATGGGAGAAATAACAAATGCAGTT	93521
QY	2165	AAGCAATGCTAAACTGFGGTTAAGGCTGATCTGAGCCCTTCTTCAATGATGTTG	2224
DB	93522	AAAGAGCTGTTGTTGTGGGTACAGGCTGATGATGAACCATCTCTTCATGCTGCTA	93581
QY	2225	CAAACTTTCCGGCATCCAAAGTTGCTGATTGGGATAGCATCAAGAATAATTTATCCCA	2284
DB	93582	CAAACTTTTAGGCAATCCAAATTAAGCTGAAACCAAGTCAAGGATTAATCCCA	93641
QY	2285	AATGGAAGACAATGATGGGATGTTGGATGAATCCAGAACCTTGGATATGGTCAGGTG	2344
DB	93642	AAAGCAGACAATGATGGGATGTTGGATGAACACACGTACGCTCAAGTATGGGCAAGTG	93701
QY	2345	TTTGTTCACTTACTGGTGGTGGACATGGAGAGTTTCTGACGATTTACATCCATTAAAT	2404
DB	93702	TTCATTGAGCTACTTCGGGTGTAATG-----	93729
QY	2405	AACAGCAGATCCACCAACAGTAATTCATTCTGAAGGGAATGTTGGTTGCAAAAAT	2464
DB	93730	-----ACAATGACAGGTTCCACCGTCAGAGGAAAAGTTGATTGCCAAAAAC	93776
QY	2465	CCATGCTTGCACTCGTGGGATATTCGTGTTTTAAAGGCTGTAAATGTTTCGAGCGCTGCAC	2524
DB	93777	CCTTGCCTCCACCACAGTGATATACGGATTCTCCATGCTGTGATGTTCTGTTTTCAC	93836
QY	2525	CACATGSPAGATTGTTGTATTCCCTCAGAAAAGGAAAAAGAC	2567
DB	93837	CACATGTTAACTGTGTGCTCTTTCCACGACAAGACCAAGCC	93879

RESULT 9

F2P3/C	F2P3	74406 bp	DNA	linear	PLN 15-SEP-1998
LOCUS					
DEFINITION	Arabidopsis thaliana BAC F2P3.				
ACCESSION	AF080120				
VERSION	AF080120.1 GI:3600045				
KEYWORDS					
SOURCE	Arabidopsis thaliana.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustoids II; Brassicales; Brassicaceae; Arabidopsids.				
REFERENCE	1 (bases 1 to 74406) Washington University Genome Sequencing Center.				
AUTHORS	The A. thaliana Genome Sequencing Project				
TITLE	Unpublished (1997)				
JOURNAL	2 (bases 1 to 74406)				
REFERENCE	Strong, C., Graves, T. and Duckels, G.				
AUTHORS	The sequence of A. thaliana F2P3				
TITLE	Unpublished (1998)				
JOURNAL	3 (bases 1 to 74406)				
REFERENCE	Waterston, R.				
AUTHORS	Direct Submission				
TITLE					

JOURNAL
COMMENT

Submitted (24-JUL-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: rwilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 3' clone is F8M12, 200 bp overlap. Actual start of this clone
is at base position 1 of F2P3; actual end is at 23552 of F8M12.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES
source

1. 74406
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="IV"
/map="unknown"
/clone="F2P3"
complement(979. .1536)
/gene="F2P3.6"
complement(979. .1536)
/note="contains similarity to Pisum sativum disease
resistance response protein 206-d (GB:U11716)"
/codon_start=1
evidence=not_experimental
/protein_id="AAC35542.1"
/db_xref="GI:3600055"
/translation="MTNQIVQVFSFLSVLLQSSVSVVPSFKLKKPKHFVLYL
HNLAYDGNANATATIVKPLGHSFGELIIINPVTLQNYLSKPVARAQGYF
YNNKTYNANVATLVFNSTKHGKTITMDANFFGLQAPDLISVGTGDFLMTGRIA
TFKTLTQGSKYFCVEMNKLKYCY"
complement(2210. .2293)
/product="tRNA-Tyr"
/note="codon recognized: UAC"
complement(2779. .3078)
/gene="F2P3.7"
complement(2779. .3078)
/note="similar to translation initiation factor IF-1"
/codon_start=1
evidence=not_experimental
/protein_id="AAC35543.1"
/db_xref="GI:3600056"
/translation="MIRCORASGGRGANRSKPAKQVKEGSKNTVIEGLVTESLPNG
MFRVDENGNIIIGYICGIRKNFRIILPGDKVKVMSYDSTKGLIIFRMSRD"
complement(4455. .10458)

tRNA

gene

CDS

gene

CDS

/gene="F2P3.8"
complement(join(4455. .5858,6701. .7012,7300. .8407,
9995. .10458))
/note="similar to several Arabidopsis thaliana disease
resistance proteins"
/codon_start=1
evidence=not_experimental
/protein_id="AAC35544.1"
/db_xref="GI:3600057"
/translation="WASSSSNRYDVPFSPRGEDVRNRFSLHLKKESEKIVTFPRD
DIKRSHTIGHELRAAIREKISVYVSENYASSWCDELTEIMKCEQGLKVPY
FYKVPDSIRKOTGKFGFLETCGCKTEERQHNRRALTDAANILGHPQWNEAY
KTTTSKDVLEKLNATPSRDNDLVGMEAHAKMESLCLSESGVIRIVGMPAGVGK
TIAIALYNOVHENENISTEVENRESYGEAGLDYGLKLHQRFVLSKLDDOKDLRY
RILGAIERLKSQKALIIDVDVNDIEQLAKAKENQWFGNKSRIYVTTNQQLVSHD
INMTQVAPSKQVLTIFCQAFKQSSPDDLKLAIEFTITLAGLTPALUKLGSFM
RKGKEWEFSFLTKSLRDLGEVYKLVGYDGLDHDHDKDLFLHACIFSGQHENYK
QMIIANNDFVSFLQVLADKSLIQKFFGRIEMHSLRQLGKEVVRKOSIYEPGKR
FUMNAKETCVLSNNTGTGVLIGSLDMCEIKELYISEKTFEEMRNVLVYLFYMSPP
IDDKMKVLQPEGLSYLPQRLHWDAYPLEPFPSPRECELVELVNSKIKLKLW
SGVQPLRLMTNMLNSNLEITLNLMEATKLRLDLGCWCELVLPSSIKNLOHLIL
LEMSCKKLEIIPTNINPSLEVHFRTCTRLQTPPEISTNIRLLNLIGTAVTPPS
VKYKIDEMERAKVRKVHVPILEKLEKLENKELETPRILYKLPQMDISY
CINIISLPKPGVSALTAVNCEISQILGHFRNKSIIHLNFINCLKUGRAQEKIHR
VIHQSYIADYLPGEHVPAYFSYRSTSIMHSKNVDLSKFNRFKVCVLYGAKIRF
EGCDIKFYKQFCKPREYVVPKHLDSPLLSKSHLCMCELFELMPPHPPTWELLHPNEF
LEVSPESRGLYKCEVKECQQLFLEPHSETFRYLSPLHYLGSGNIGNSSSIEIIF
VDOESSSDSEBIIYADQESSSGIEEIIHAEREGTNRKRKSVMRWIKYVGARKMGLSLE
CLKPWTR"
complement(14859. .17002)
/gene="F2P3.9"
complement(join(14859. .15293,15428. .15574,15656. .16209,
16261. .16801,16868. .17002))
/gene="F2P3.9"
/note="similar to initiation factor IF2-beta (Pfam:
GTP_BFTU.hmm, score: 226.67)"
/codon_start=1
evidence=not_experimental
/protein_id="AAC35536.1"
/db_xref="GI:3600049"
/translation="MTLLELSKRTGESVAVLQSLINVGTFSSFEFTISVDVAELLA
MEIGINVRROHSTGESEILPRPPVTVMGVDHDKTSLDLALRNTVAARAGGITOH
VGAFVGMPSDSTSEITLDPGHAASEMRARGAATDVIVLVAADQVMPQLEAI
AHARSANPVVAINKDKPGANPEKVIKLTISEGIELEDIGGNVQVNDKNGFGLIN
FLQFGLAKSLGLKLEALILLQAVMDLKAQVDPQAVVVEARLDKRGPLATIIY
KAGTLVRGQHVVGQWGLRAIRDMIGKTTDRATPAMPVEIEGLKLPMDAGDDVIY
ESEERARMLSEGRKRYEKRLLEAEARELEAEARLETESEEGFVRYELPIVYKSDY
QGTAAVADALRTLNSPOVSNIVHSGVGAISHDLDAQACGACIVGFNVKGGSTGN
LSAAGSVKVPFHHVYIHLLEDIGNLIVEKAPGVSEMEVSGEAEVLSIFKILGRKTE
EDGVNAGCKVMDGRVCBSGLMRLRSCEVVEGSCASLKREKQDVEQVKGNEGGLV
FGDWNDFRGVDVQCMPEVIRKPKFISSEANVREC"
complement(19889. .21494)
/gene="F2P3.10"
complement(join(19889. .19924,20035. .20134,20239. .20435,
20635. .20791,21177. .21277,21393. .21494))
/note="similar to vacuolar ATPases; coded for by A.
thaliana cDNA T43489; coded for by A. thaliana cDNA
T04528; coded for by A. thaliana cDNA R64882; coded for by
A. thaliana cDNA T46025; coded for by A. thaliana cDNA
RA042676; coded for by A. thaliana cDNA T21569; coded for
by A. thaliana cDNA T45327; coded for by A. thaliana cDNA
RA042335; coded for by A. thaliana cDNA W43660"
/codon_start=1
/protein_id="AAC35545.1"
/db_xref="GI:3600058"
/translation="MNDGVSQRQIQMVRFIRQEAEEKANEISVSAEEFNEIKQLV
EAEKKIRQDYKEKQADVRKIKIDYSQMLNARIKVLQAOQDDIVNMQDQAADLLN
VSRDYAYKQLLKDLIVCLRLKPEPSVLLRCREBGLVEAVLDDAKEEYAGAKVH
APEAVADTKIFLPPPKSNDPFLGHCSSGVVLAASRDKIVCENTILDAFLVAFRMKLP
VIRKSLFGQVTA"
25791. .26654

gene

CDS

gene

CDS

gene

```
CDS
/gene="F2P3.5"
25791..26654
/gene="F2P3.5"
/note="Contains similarity to AP2 domain containing
proteins"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC35537.1"
/db_xref="GI:3600050"
/translat="METEKVSLPRILISVTPYATDSSDEEVEYDFDALSTKRER
VKYKVEWLVDSYDEKPKMKRKRVTVPVVTATRKRGVQRQWKAHEI
RDPKRRVWLGTEDEAEAAIVYDAAIQLRGNALNFPVPPVBNBEEATVNG
VDFIIIGVCCVSPVSEVSTAYKEEFVGVSTAEIIVVKKEPSNGDFPAP
LFSDVFGSTMSSESGDGLFDNLFADMSGSGFGSGSGFSSHWVHEDFDIG
DLFGSDPLTV"
gene
complement(37194..41062)
/gene="F2P3.11"
complement(join(37194..37985,38074..38224,38340..40246,
40511..41062))
/gene="F2P3.11"
/note="similar to hypothetical proteins in
Schizosaccharomyces pombe (GB:298533) and C. elegans
(GB:248334 and 278419)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC35535.1"
/db_xref="GI:3600048"
/translat="MVSETTNRSTVKISNPOTIVADELLRELEHLGSDTVFALEI
PTDRNKKPDRFAVQPTTLEKVSQAOLLSQKLFKTHNLRLSEAYDDIIIPVDP
RKRLDDVLIVTGVPPSEDEKFALEKWDGVRVILTEKRVFVWWSGDCYKIEVRF
EDITETLSCVGNPASBIDAFLLKLYGPKVRAVTHIATKFSDRYRCKEDPFM
WIRITDFSGKSIGTSCFLEHNGSTMIDIESGLPYREDILSLTVYDQKTFASNA
QIVPLNNAIILGLPEYFELPOLNALYVHAQKISLFAASDMELIKILRMSLEIALVIL
```

```
Query Match      8.4%; Score 314.6; DB 8; Length 74406;
Best Local Similarity 55.2%; Pred. No. 4e-54;
Matches 689; Conservative 0; Mismatches 539; Indels 21; Gaps 3;
```

```
QY 1277 GTAACACCATGCAAGTTATTGTTGCTCCAGAGGTTAATGTTCCAAAGCGGTTCTC 1336
DB 39643 GTTACACCTCAAGATCTATCTGTTGGTCCAGAGCTTGAGACTGCAAAATATGTTGTG 39584
QY 1337 CGCAATTATTCGAACATAGATAACTTCTCTGTTTCTTTGTTGATGAGAGTGG 1396
DB 39583 AAGAACTTTCGAGAGCATGCTCCGATTTCTATGAGTACTTTTGTGGAAGAGATTGG 39524
QY 1397 GAGAACTGTATTCACAGACTTATACCAAGCAAGTACTGAGTGGTGTGCTAGG--- 1453
DB 39523 AGCAAGCTTCCCGCAAAATGCTCTCTGTGAACCTCAAGGAAGGCTATTTTGTGAAGCCC 39464
QY 1454 -----ACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTTGTAATTGGT 1507
DB 39463 TCTAGAACCAACATTTATACAGGCTGTGTCATCTTGGAGAGGATACCGTTGGG 39404
QY 1508 GATAAAAATTTGAATTTCTTCATTCATCGAGCCAGTTCGGGGAATATCATGATGG 1567
DB 39403 CCTAAGAGTTTGAAGTTCTTCTTGGCTTTTTCAGCGCAACTCGAGGAATATCGGTCG 39344
QY 1568 ATCTTGCATCAAGACCTGGCTTACTGCAATGATATAAGAGCTTGGATGGTGATTTT 1627
DB 39343 ATGTTTGTCTTAACGAGAAATTAAGCGGAAGATATAAGAGATGATGGTGGTTGTTTC 39284
QY 1628 TCGCAGATCAAGAAATGTCGCAAAATATGTCGCAGACTTGGTCAATCTTTTGGTTCCTCC 1687
DB 39283 CGTAAATCCGAGCATTTCCAAATGCTGCTAGTAGTGGTTCAGTTCTTCTAGTCTTCG 39224
QY 1688 AGAGAGACTTTGAGTGTCTTTAGGCAATGAGATTGAAGTTATTCCTCCGATGTAAGGT---- 1743
DB 39223 CGGCAAACTTATTGTCGTCGACAAAGATGTTGGAGCAGATTCCTCCGATCGAAGTGACA 39164
QY 1744 --TCATGGAACAGCTATGCTTTCTGATGGAATGTTAAATATCTGGTACTTTGCT 1801
DB 39163 ACTGATGGTGGTACTGCTCTCGGATGCGATTTGGAAATTTCACTTGCATTTGCC 39104
```

```
QY 1802 CATAGAGTTGCCCTCAAAATGTCGCTTCAATATACCCCATCTGCTTCCAGATTCGTTAT 1861
DB 39103 AGCAAGTTGCACAGAGTGTGGATTGAGTCATGCTCCCTTCTCCCTTCAAAATCGATAC 39044
QY 1862 GGTGGATATAAAGTGTGTGGTGTGATCCGGAATCATCAATGAAGTTGTCTTTGAGA 1921
DB 39043 GGTGGCTACAAAGGTGTGATTGCTGTTGACCGGAGTTCTCTCCGAAAATGTTCTCGCT 38984
QY 1922 AAGAGCATGTCGAAATGAATCAGACAAACATRAAGTTAGATGCTCTGATGAGCAGAAA 1981
DB 39863 GATAGTATGCTTAAATTTGATCTGCAACACAGGATGCTGAACGTTACCGAGTGGACAG 38924
QY 1982 TATCAGCTTGTATTCTTAATCTCAACTGATTACGCTCTTGTCTACACTTGGAGTAAA 2041
DB 38923 TCGATGCTTGTCTTTAAACCGGAGATCATTTGCTCTTTGTCGACCTTGGAAATGAA 38864
QY 2042 GATGAAGTTCTGACAGAGAAAGAGAGCTGTAGATCAGCTGTGATCTATCTTGCAT 2101
DB 38863 GATGCAATGTTGAGCGCATGCAAGCGGTGCATTTATCTATGCTTGGGAATATGCTTGAA 38804
QY 2102 GATTCTTTGAAGGCACAGAGGCTTTTGAATTTGATGCTCTCTGAGAGAAACACATAATT 2161
DB 38803 GACCGCATGCGACACTTAATGTTCTCCAAATTTAGCGGAGAAATTTCCAAAGATTG 38744
QY 2162 CTCAAGCAATGCTAAACTGTGTTATATAAGCTGATGCTGAGCCCTTTCTTTCAATGATG 2221
DB 38743 CTAGTTAAGATGCTGCTTCAAGGATATGACCGAGTTTCAAGACCTTACCTCTCAATGATG 38684
QY 2222 TTGCAAAACCTTCCGCGCATCCAAAGTTCTCTCGATTTGGGACTAGATCAAGAATATTATT 2281
DB 38683 CTTCGTGTGCACACAGAGAGCCAGCTTCTGAACTAAAGAGACAGATGCAGATATTGTA 38624
QY 2282 CCAATGGAAGAACATGATGGGATGTTGGATGAATCCAGAACCTTGGAAATATGTCAG 2341
DB 38623 CCGAAAGAGCGGATCTTGAAGTGTGATGATGAAATGGTATCTCTGAGTATGGCAA 38564
QY 2342 GTGTTTGTTCAGTTTACTGCTGTCGTCGATGAGAGGTTTCTGACGATTTTACATCCATT 2401
DB 38563 GTGATGTTCTGTGAACCTGACTAAAGCGGAATGAAATCTCCGATCAGAGCTACTTT 38504
QY 2402 AATAACAGAGATCCCAACACAGTAATTTCTTCTGAAGGAAATGTTGTTGTTGCAAAA 2461
DB 38503 -----CGCAAGATTTGATGAGGAACATCTGTGTTATTGGGAAGTGGTGGTACGAAA 38450
QY 2462 AATCCATGCTGCTGCTGCTGATATCTGTTTAAAGGCTGTAAATG 2510
DB 38449 AACCATGCTTCTACCTGGAGATATTAGAGTTCTTGTATGCTATATG 38401
```

```
RESULT 10
ATT22B4 108598 bp DNA linear PLN 10-MAR-2000
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone T22B4 (ESSA)
DEFINITION Arabidopsis thaliana.
ACCESSION AL049876
VERSION AL049876.1 GI:4850281
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE 1 (bases 1 to 108598)
AUTHORS Bevan, M., Pohl, T., Weizenegger, T., Bancroft, I., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 2 (bases 103489 to 103876)
AUTHORS Bevan, M., Volckaert, G., Grymonprez, B., Voet, M., Robben, J., Bancroft, I., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 108598)
AUTHORS EU Arabidopsis sequencing, project.
```


TITLE	Direct Submission		/gene="AT4g11030"
JOURNAL	Submitted (09-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk	intron	/number=5 11114..11201 /gene="AT4g11030" /number=5 11202..11278 /gene="AT4g11030" /number=6 11279..11374 /gene="AT4g11030" /number=6 11375..11414 /gene="AT4g11030" /number=7 11415..11499 /gene="AT4g11030" /number=7 11500..11594 /gene="AT4g11030" /number=8 11595..11678 /gene="AT4g11030" /number=8 11679..11846 /gene="AT4g11030" /number=9 11847..11932 /gene="AT4g11030" /number=9 11933..12043 /gene="AT4g11030" /number=10 12044..12129 /gene="AT4g11030" /number=10 12130..12217 /gene="AT4g11030" /number=11 12218..12334 /gene="AT4g11030" /number=11 12335..12401 /gene="AT4g11030" /number=12 12402..12474 /gene="AT4g11030" /number=12 12475..12550 /gene="AT4g11030" /number=13 12551..12624 /gene="AT4g11030" /number=13 12625..12742 /gene="AT4g11030" /number=14 12743..12814 /gene="AT4g11030" /number=14 12815..13060 /gene="AT4g11030" /number=15 13061..13166 /gene="AT4g11030" /number=15 13167..13306 /gene="AT4g11030" /number=16 13307..13486 /gene="AT4g11030" /number=16 13487..13669 /gene="AT4g11030"
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ .	exon	
FEATURES	Location/Qualifiers	intron	
source	1..108598 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4"	exon	
misc_feature	1..11221 /note="overlap to BAC F25124; please refer to this entry for analysis and annotation"	intron	
gene	8990..13999 /gene="AT4g11030"	exon	
CDS	join(9890..10044,10569..10632,10733..10822,10923..10979,11065..11113,11202..11278,11375..11414,11500..11594,11679..11846,11933..12043,12130..12217,12335..12401,12475..12550,12625..12742,12815..13060,13167..13306,13487..13669,13742..13861,13943..13999) /gene="AT4g11030"	intron	
	/note="strong similarity to acyl-CoA synthetase - Brassica napus (rape), FID:e217854 Contains Putative AMP-binding domain signature AA228-239 contains EST gb:A1992650.1, N65639, T20845, T43231, N38362, H77181, H76835, A1999263.1, T45466, AA395246" /codon_start=1 /product="putative acyl-CoA synthetase" /protein_id="CAB43038.1" /db_xref="GI:4850282"	exon	
	/translation="MFSQKRFIEVEAAKEATDGNPSVGPVYRSTFAQGFPPIDIGI QSCWDIFRTAVERYPNNRMLGRREISNGKAGYVWKTYEYDIVIKLGNLSRSCGK EGECGDIYGINCCWEIISMEACNAHGLYCVPLDYTLGAGAVEFIISHAEVSIAPVEEK KIPLEFVCNPTKMYKTIVVSGVKGPEQKEAEKGLVTHSWDFLKLGEKQYELP IKPSPDICTIMYSGTGDPKGVNISNESIVTITGVHVELGNVNASLSEKQVLSYL PLAVFRATEEELIQVGSIGFWRGDVKLLIEDLGLKPSIECAVPRVLDVRYTGLO QKLSGGGFFKKKVFDAFYSYFGNMKQSHVAASPFCDKLVFNKQGLGGNVRIL SGAAPLASHIESLRVAVACCNVLOGYGLTESCAGTATFPDELDMLGTGVPVPVNI RLESVPMDYDALGTPRGECIRGKTLFSYKREDTLKEVFDGLWHTGTVGEVQD NGSKIIDRKNIFKLAQGYVALENVYSQVEVIESIWIYVNSFESFLVAIANPA QOTLERAVGVNGVDNFNSIQNAKAKAFILGLVKTAKENKLGFEILIKDVHLEPVA FQDERDLITPTKKRPQLLKYYQNVIHENYKTKESLASGQ"	intron	
	8990..10044 /gene="AT4g11030"	exon	
	/number=1 10045..10568 /gene="AT4g11030"	intron	
	/number=1 10569..10632 /gene="AT4g11030"	exon	
	/number=2 10633..10732 /gene="AT4g11030"	intron	
	/number=2 10733..10822 /gene="AT4g11030"	exon	
	/number=3 10823..10922 /gene="AT4g11030"	intron	
	/number=3 10923..10979 /gene="AT4g11030"	exon	
	/number=4 10980..11064 /gene="AT4g11030"	intron	
	/number=4 11065..11113	exon	

/gene="AT4g11030"
/number=5
11114..11201
/gene="AT4g11030"
/number=5
11202..11278
/gene="AT4g11030"
/number=6
11279..11374
/gene="AT4g11030"
/number=6
11375..11414
/gene="AT4g11030"
/number=7
11415..11499
/gene="AT4g11030"
/number=7
11500..11594
/gene="AT4g11030"
/number=8
11595..11678
/gene="AT4g11030"
/number=8
11679..11846
/gene="AT4g11030"
/number=9
11847..11932
/gene="AT4g11030"
/number=9
11933..12043
/gene="AT4g11030"
/number=10
12044..12129
/gene="AT4g11030"
/number=10
12130..12217
/gene="AT4g11030"
/number=11
12218..12334
/gene="AT4g11030"
/number=11
12335..12401
/gene="AT4g11030"
/number=12
12402..12474
/gene="AT4g11030"
/number=12
12475..12550
/gene="AT4g11030"
/number=13
12551..12624
/gene="AT4g11030"
/number=13
12625..12742
/gene="AT4g11030"
/number=14
12743..12814
/gene="AT4g11030"
/number=14
12815..13060
/gene="AT4g11030"
/number=15
13061..13166
/gene="AT4g11030"
/number=15
13167..13306
/gene="AT4g11030"
/number=16
13307..13486
/gene="AT4g11030"
/number=16
13487..13669
/gene="AT4g11030"

/number=17
13670..13741
/gene="AT4g11030"
/number=17
13742..13861
/gene="AT4g11030"
/number=18
13862..13942
/gene="AT4g11030"
/number=18
13943..13999
/gene="AT4g11030"
/number=19
16931..18437
/gene="AT4g11040"
Join(16931..17069,17349..17821,17991..18142,18314..18437)
/gene="AT4g11040"
/note="similarity to phosphoprotein phosphatase (EC 3.1.3.16) 2C Arabidopsis thaliana, PIR2:S55457"
/codon_start=1
/product="putative protein"
/protein_id="CAB43039.1"
/db_xref="GI:4850283"
/translation="MKTDLTLPIDAEKDGSKRVVADSGYVGGDRPVKLPKIE
NNGDVTSFGTHVLDALMAEVALDKDKNAGHGVSVMGVRAMTAVSTVVDIEI
PSVDIFGIDGLAKFEFEDRLRLVKESKACHGCVAAADKNKMKSCFSAVGTG
TTTSAVTVIGKEVIVLCRGARVVLISHDGVALPLCHHHKDKGVEGILLKHKK
IDDFVLACDGLMDVYDDDTYQLVRKCLYGLPPDDCCISSESSTKAIVILAEIAR
GSXENINVIDLKSSTVS"
16931..17069
/gene="AT4g11040"
/number=1
17070..17348

Query Match 8.4%; Score 314.6; DB 8; Length 108598;
Best Local Similarity 55.2%; Pred. No. 3.9e-64;
Matches 689; Conservative 0; Mismatches 539; Indels 21; Gaps 3;

QY 1277 GTACACCATCCAAAGTTATTTTGTGGTCCAGAGGTTAATGTTTCCAAATCGGGTCTC 1336
DB 53711 GTTACACCTCAAGATCTATCTGTTGGTCCAGAGCTTGAGCTGCAATATATGTTG 53770
QY 1337 CGCAATATTCGAAGACATAGATACTTCTGCTGTTCTTTTGTGATGAGAGGTGG 1396
DB 53771 AAGAACTTTGCAGAGCATGCTCGGATTCATGAGAGTTACTTTTGTGGAAGAGATTGG 53830
QY 1397 GAGAACTGTTTCTACAGACTTATACCAAGCAAGTACTGCAAGTGTGTCAGG--- 1453
DB 53831 AGCAAGCTTCCGCAATGCTCTCTGTGAATCCAGGAGGCTATTTTGTGAAGCCC 53890
QY 1454 -----ACAAACATCTATGAGAGATCTTATCAACTCTCGCGAAGGCTTTGTAATTGGT 1507
DB 53891 TCTAGAACCACTTATAACAGGGTGTGTCAATCTTGGAGAAGGATACCGTTGGG 53950
QY 1508 GATAAAATTTGAATTTCTTTCATTCGAGCCGAGTTGCGGGATAATTCAGTGTGG 1567
DB 53951 CTAAGAGGTTTGAATTTCTGCTTTTTCAGCGAGCCAACTCGAGGAAATTCGGTCTGG 54010
QY 1568 ATGTTGTCATCAAGACCTGGCTTACTGCAATGATATAAGAGCTTGGATGGTGATTTT 1627
DB 54011 ATGTTTGTCTTCAAGGAAAGTAAACCGGAGATATAAGAGATGATGATGGTGTTC 54070
QY 1628 TCGCAGATCAAGAAATGTCGCAAAATATGCTCCAGACTTGTGTCATCTTTTGGTTCCTCC 1687
DB 54071 CGTAAATCCGAAGCATTTCCAAATGTGCTGTAGGATGGTGTGCTTTCAGTGTCTCG 54130
QY 1688 AGAGAGACTTTGAGTGTCTTAGGATGAGATTGAAGTTATCCCGATGTAAGGT---- 1743
DB 54131 CGGCAAAACACTTATTGTCGTCACAAAGATGTGGAGCAGAGATCCCGACATCGAAGTGACA 54190
QY 1744 --TCATGGAAACAGCATGCTCTTTTCTGTATGGAATTTGTTAAATATATCTGTGACTTTGCT 1801
DB 54191 ACTGATGGTCTGATCTACTGCTTCTCGGATGSCATTTGGAAAATTTTCACTTGCATTTGCC 54250

QY 1802 CATAGAGTTGCTCAAAATGTGGCCTTCAATATACCCCATCTGCTTCCAGATTCGTAT 1861
DB 54251 AAGCAAGTTGCACAGAGTGTGGATTGAGTCACTGTCCTTCTGCTTTCAATATCGATAC 54310
QY 1862 GTGGATATAAAGGTGTTGTGGTGTTCATCCGGATTCATCAATCAAGATTCTCTTTGAGA 1921
DB 54311 GTGGCTACAAAGTGTGATTGCTGTGACCGCAGTTCCTCCGAAAGTGTCTCTGCGT 54370
QY 1922 AAGAGCATGTCGAAATATGAATCAGACACATATAAGTTAGATGTCCTTGGATGGAGCAA 1981
DB 54371 GATAGTATGCTTAAATTTGACTCGAACACAGGATGCTGAACGTTACAGGTGGACAGAG 54430
QY 1982 FATCAGCCTTGTATCTTAATCGTCAACTGATTACGCTCTTGTCTACACTTTGGAGTGAA 2041
DB 54431 TCGATGCCCTTGTCTTAAACCGGAGATCATTTGCCCTTTGTGACCCCTTGGATAGAA 54490
QY 2042 GATGAAGTTTCTCGAAGACAGAAAGAGAGCTGTAGATCAGCTTGAATCTTCTTGTGAT 2101
DB 54491 GATGCAATCTTTGAGCGGATGCAAGCGGTGCATTTATCTATCTTGGGAATATGCTTGA 54550
QY 2102 GATCTTTGAAGGCACAGAGGCTTTGGAATGATGCTCTCTCGGAGAGAACACTAATATT 2161
DB 54551 GACCGGATGCAGCAGCTTAAATGTTCTGCAAAATGAGCGGAGAAAAATCCAAAGATTG 54610
QY 2162 CTCAAGGCAATGCTAAACTGTGGTTATTAAGCCTGATGCTGAGCCCTTCTTTTCAATGATG 2221
DB 54611 CTAGTTAAGATGCTGCTTCAAGGATATGACCGAGTTCAGAACCTTACCTCTCAATGATG 54670
QY 2222 TTGCAAACTTCCGCGCATCCAGTTGCTGATTTCCGAGCTAGATCAAGAAATATTATT 2281
DB 54671 CTTCGTGTGCACCGAGAGCAGCTTCTGAACTAAAGCAGAGTGCAGGATCTTGTGA 54730
QY 2282 CCAATGGAAGACAATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCAG 2341
DB 54731 CCGAAGAGCGATCTTGATCGTTCATGATGAATGGTATCTCTGGAGTATGCCAA 54790
QY 2342 GTGTTTGTGAGTTTACTGGTGTGACATGAGAGATTTCCTGACGATTTACATCCATT 2401
DB 54791 GTGTATGTTCTGTAACCTGACTAAAGCGGAATGAAATCTCGCATCAGACGCTACTT 54850
QY 2402 AATAACAGCAGATCCACCAACAGTAAATTTCTTCAAGGGAATGTGGTTGTGCAAAA 2461
DB 54851 -----CGCAAGATTGATGAGGAACATCTGTGTTATTGGGAAAGTGTGCTGAGGAAA 54904
QY 2462 AATCCATGCTTGCATCCTGGTGATATCGTGTTTTTAAAGCTGTAATG 2510
DB 54905 AACCCATGCTTCCACCTGGAGATATTAGAGTTCTTGTGCTATATATG 54953

RESULT 11
ATCHRIV31 198301 bp DNA linear PLN 16-MAR-2000
LOCUS Arabidopsis thaliana DNA chromosome 4, contig fragment No. 31.
DEFINITION
ACCESSION AL161531
VERSION AL161531.2 GI:7267806
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 81065)
Fohl, T., Weizenegger, T., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
Unpublished
2 (bases 73007 to 171954)
Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W.,
Lemcke, K. and Mayer, K.F.X.
Unpublished
3 (bases 75956 to 76343)
Volckaert, G., Grymonprez, B., Voet, M., Robben, J., Mewes, H.W.,
Lemcke, K. and Mayer, K.F.X.
Unpublished
JOURNAL


```

/ gene="AT4g11100"
/ number=1
13862..13902
/ gene="AT4g11100"
/ number=1
13903..14213
/ gene="AT4g11100"
/ number=2
16403..20912
/ gene="AT4g11110"
join(16403..16547,16590..17213,17271..17866,18203..18321,
18736..19568,19624..19859,19922..20239,20334..20393,
20721..20912)
/ gene="AT4g11110"
/ note="EST GB:A404849 spans an intron in 3' nontranslated
region and marks 3' end
similarity to photomorphogenesis repressor COP1 -
Arabidopsis thaliana, F1R2:10112
Contains Prokaryotic membrane lipoprotein lipid attachment
site AA405-415; Trp-Asp (WD-40) repeats signature
AA799-813; Trp-Asp (WD-40) repeats signature AA919-933
contains EST gb:A199586.1, T46420, AA404849"
/ codon_start=1
/ product="COP1 like protein"
/ protein_id="CAB81212.1"
/ db_xref="GI:7267810"
/ translation="MDEGSGVDYSRIDEADVAHLQFNSEQSPKFNIEVREYKQV
OREADGKNGDGVLELIGNEPCSSRQNTDEGDVVELTVKTCGSSMAIYGRPSRAR
LEMNRSDFLHRELDGLPGSSSKKVIDRGTVSILRNAGKMSLPETSNQCAIIV
NGEANHLTVNRPVPVEALSHGKIKMLSGSFQFVRKTLGKGVTFRPPNN
RSKARNDDQTVASSGSALVIANISAKISSIPLAAYDDTHRCGGEGSLRKLKSE
RQVNEKAEKIFQIVDHDVCSHQGVILCDLRPSFKIFKFNANVYVSGSQRSF
DSNNKETLSQLENPLVRLRGDTSSLSIPAKKSGSPSSRQMPFQAGGVNIOTE
Query Match      8.4%; Score 314.6; DB 8; Length 198301;
Best Local Similarity 55.2%; Pred. No. 3.6e-64;
Matches 589; Conservative 0; Mismatches 539; Indels 21; Gaps 3;

QY 1277 GTAACACGATGAAAGTTATTTTGTGGTCCAGAGGTTAATGTTCCAAATCGGGTCTC 1336
DB 26178 GTTACACCTCAAGATCTATCTGTGGTCCAGAGCTTGAGACTGCAAAATGTTGTG 26337
QY 1337 CGCAATATTCGAGACATAGATAACTTCTCGTGTCTTTTGTGATGAGAGTGG 1396
DB 26238 AGAAGCTTGCAGACATGCTCGGATTCATGAGATTAAGTTTGTGGAAGAGATTGG 26297
QY 1397 GAGAACTGTATTCACAGACTTATACCAAGCAAGTACTGGAAGTGGTCTAGG--- 1453
DB 26298 AGCAAGCTTCCCGCAAAATGCTCTCTGTGACTCCAAAGAGGCTATTTTGTGAAGCCC 26357
QY 1454 -----ACAAACATCTATGAGAGGATCTATCAACTCTCGGAAAGGCTTTGTAATTGGT 1507
DB 26358 TCTAGAACCAACATTTATACAGGTTGTGTCATCTTGGAGAGGATTAACGTTGGG 26417
QY 1508 GATAAAAAATTTGAATTTTTCATTTTCATCGAGCCAGTTGCGGGAATTAATTCAGTGG 1567
DB 26418 CTTAAGAGTTTGTAGTCTCTGTGCTTTTTCAGGAGCCAACTCGAGGAAATTCGCTCG 26477
QY 1568 ATGTTTGCATCAAGACCTTCTACTGCAATGATATAAGAGCTTGGATGGTGATTTT 1627
DB 26478 ATGTTTGTCTTACAGGAAAGTAAAGCGGAAGATATAAGAGATGATGGTGGTTGTTT 26537
QY 1628 TCGCAGATCAAGATGTCGCAAAATATGTCGACACTTGGTCAATCTTTTGGTTCCTCC 1687
DB 26538 CGTAAATCCGAAGCATTTCCAAATGCTGCTAGAGTGGTCAAGTTTCAAGTCTTCG 26597
QY 1688 AGAGAGCTTTGAGTGTCTTTAGCGATGAGATGTAAGTTATTCGCGATGAAGGT---- 1743
DB 26598 CGGCAAACTTATTTCTCGGTGCACAAAGATGTGGAGCAGATTCGCGACATCGAAGTGACA 26657
QY 1744 --TCATGGACACGATCTCTTTCTGATGGAATGGTAAATATCTGTCACATTTGCT 1801
DB 26658 ACTGATGGTGGCTGATTTACTGCTTCTCGGATGGCATTTGGGAAATTTCACTTGCATTTGCC 26717

```

```

QY 1802 CATAGAGTTGCCTCAAAATGTGGCCTTCAATATACCCCATCTGCTTTCCAGATTCGTTAT 1861
DB 26718 AACCAAGTTGCACAGAGTGTGGATTGAGTCATGTCCTTCCTTTCAAAATCGATAC 26777
QY 1862 GTTGGATATAAAGTGTGTGGGTGTGTGATCCGGATTCATCAATGAAGTTGCTTTGAGA 1921
DB 26778 GTTGGCTACAAAGTGTGATGTGTGTGACCGGAGTTCCTTCGGAAGAGTGTCTCTCGCT 26837
QY 1922 AAGAGCATGTCGAAATGAATGAATCAGACACATAAAGTTAGATGTCCTTGGATGGAGCAAA 1981
DB 26838 GATAGTATGCTTAAATTTGACTCGACACACAGAGATGCTGAACGTTACCAAGTGGACAGAG 26897
QY 1982 TATCAGCCTTGTATCTTAACTGCTCAACTGATPACCGCTCTTGTCTACACTTGGAGTGAAA 2041
DB 26898 TCGATGCTTGTCTTCTTAAACCGGAGATCATTTGCTTGTCTGACCCCTTGGAAATAGAA 26957
QY 2042 GATCAAGTTCTCAACAGAGCAAAAGGAGCTGTAGATCAGCTTGTGATCTTCTTTCGAT 2101
DB 26958 GATGCAATGTTGAGCGGATGCAAGGGTGCAATTTATCATGCTTGGGATATGCTTGAA 27017
QY 2102 GATTTCTTTAAGGCACAGGAGGCTTTTGAATGATGTCCTCTGGAGAGAACATATATT 2161
DB 27018 GACCGGATGCAGCAGCTTAATGTTCTGCAGAAATTTGACGCGAGAGAAATTCCAAGAAATTG 27077
QY 2162 CTCAGGCAATGCTAATCTGTTGTTTAAAGCTGATGCTGAGCCCTTTCTTTCATGATG 2221
DB 27078 CTAGTTAAGATGCTGCTTCAAGGATATGACCGAGTTTCAAGCTTACCTCTCAATGATG 27137
QY 2222 TTGCAAAACCTTCCGGCGCATCCAAAGTTGCTCGATTTCGGGACTAGATCAAGAATATTTAT 2281
DB 27138 CTTCGTGTGCACACAGAGAGCCAGCTTCTGAACTAAAGAGCAGATGCAGGACTACTGTA 27197
QY 2282 CCAATGGAAGACATGATGGATGTTTGGATGAATCCAGAACCTTGGATATGGTCAG 2341
DB 27198 CCGAAAGAGGAGTCTTGTGCTGATGATGATGAAATGGTATPCCTGGAGTATGGCAA 27257
QY 2342 GTGTTTGTTCAGTTTACTGCTGCTGGACATGGAGAGTTTCTGACGATTTTACATCCATTT 2401
DB 27258 GTGATGTTCTGTAAACCTGACTTAAGCGGAATGAAATCTCGCATCAGAGCTACTTT 27317
QY 2402 AATAACAGCAGATCCCAACACAGTAATTTCAATCTTGAAGGGAATGTTGTTGTCAGAAA 2461
DB 27318 -----CGCAAGATTTGATGAGGAACATCTGTGTTTATGGAAGTGGTGGTGCAGCAA 27371
QY 2462 AATCCATGCTTGCATCTGCTGATATTCGTGTTTATAAGGCTGTAAATG 2510
DB 27372 AACCAATGCTTCCCTCGAGATTTAGAGTCTTCTGATGCTATATG 27420

```

RESULT 12
OSJN00104 121492 bp DNA linear PLN 17-JUL-2002
LOCUS OSJN00104
DEFINITION Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBb0048502,
complete sequence.
ACCESSION AL606653
VERSION AL606653.2 GI:21741220
KEYWORDS HTG.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
AUTHORS
Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q., Hu, X.,
Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Weng, Q. J., Zhang, L.,
Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Zhang, Y. J., Lu, Y., Li, C.,
Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K., Zhou, B.,
Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P., Fu, G.,
Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F., Tu, Y. F.,
Jia, J., Yin, H. F., Zhang, Y., Cai, Z., Chen, J., Kang, H., Chen, X. Y.,
Shao, C. Y., Sun, Y., Hu, Q. P., Zhang, X. L., Zhang, W., Wang, J. J., and
Ding, C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L., Zhu, F. H. and

TITLE Direct Submission
 Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,
 Chinese Academy of Sciences, 500# Cao Bao Road, Shanghai 200233,
 CHINA. E-mail enquiries: bhanncgr.ac.cn. Clone requests:
 bhanncgr.ac.cn
 REMARK Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
 clone: OSJNB0048E02.
 COMMENT On Jul 12, 2002 this sequence version replaced gi:15594110.
 Web site: <http://www.ncgr.ac.cn>
 ----- Summary Statistics
 Assembly program: Phrap

NOTE: This is a complete sequence.
 Genes were identified by a combination of several methods: Gene
 prediction programs including Fgenesh (<http://www.softberry.com/>),
 gensean (<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkESM
 (<http://genemark.biology.gatech.edu/GeneMark/>), tRNAscan-SE (Sean
 Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the
 complete sequence against NCBI none redundant protein database (nr)
 (<ftp://ncbi.nlm.nih.gov/blast/db>) and the EST database at NCGR.

FEATURES

Location/Qualifiers
 1..121492
 /organism="Oryza sativa"
 /variety="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="4"
 /clone="OSJNB0048E02"
 /clone_lib="CUGI-OSJNB"
 BASE COUNT 34036 a 25389 c 25204 g 36813 t 50 others
 ORIGIN

Query Match 8.4%; Score 314.2; DB 8; Length 121492;
 Best Local Similarity 52.8%; Pred. No. 4.7e-64;
 Matches 865; Conservative 0; Mismatches 733; Indels 41; Gaps 7;
 QY 953 TCGGCTCTGGTTCCTCAAGTCCAGCTCCGGAAGGAATTCCTCAAGATTG 1012
 DB 99416 TCCATGTGGTCCCTTGTTCATGTCCGAGGACTATTCAGTCCCTTATGAAGTCTG 99475
 QY 1013 TTCAAATAGTCTTTGGTACAGATGGATGCATACCTGGCGCAGCAATTAATGTCTAC 1072
 DB 99476 TTTCGTCTGAATCTCTCATGCATATGGGGAAGATAGTCGCAAGCATGTTAATGCCGAT 99535
 QY 1073 TTTTCCGATAGTATCTCCGAGGAGAAATGTGGCATGCATGAGCATGCCCTTAGAG 1132
 DB 99536 CTGTTTAAAGCCCTTCAAGACTACCGGTTGATGTCCTCAAG-----GAGAAATTTTGA 99589
 QY 1133 AAACGTACTATATAAGAGTGTCTGTTATGATCCCGTGGTGGCTCCTACAGCAG--- 1189
 DB 99590 AAATGCACAAATAAAGTACCTGTACGGGCTTACAAATTTATCCACAGGAGCT 99649
 QY 1190 --TATGATGGGTATCTCAAGGGTAGACAACTCCAAATCTCCGTCCTACATTAGATG 1247
 DB 99650 TATAGCATGAAGAGAGGACCAATAATTTTGTCTCCCAATGAAGGTGAAGGAGTGA 99709
 QY 1248 ATGGGTGTGTATGTAAGAAGGGTCTTAGTAACACACCATGCAAAAGTTTATTTTGTGTC 1307
 DB 99710 GAAATGATGAAGTGTATAGAGTTACATACTCCCTCCAAATATTCGTGTTGAC 99769
 QY 1308 CAGAGGTAAATGTTCCAAATCGGTTCTCCGCAATTAATTTGAAGACATAGATACTTC 1367
 DB 99770 CTGAAGAAGAGGTTACAAATATATGTGGTTAAGCATCATCTGTCATATGCTTCTGAC 99829
 QY 1368 TTCGTGTTCTTTCTGATGAGGAGTGGGAAACTGTATCTA-----CAGACT 1418
 DB 99830 TGAGAGTTACTCTCTGATGAGATTGAGTAAGCTTCTCTAATGCAATCTCAGCTA 99889
 QY 1419 TATTACCAAAAGCAAGTACTGGAAGTGGTGTGAGACAAACATCTATGAGAGGATCTTAT 1478
 DB 99890 GAATTGAACAAGGTTCTCTTTCTTAAGCCCTTTCAAACTGGTCTATATTATCGGATCTCT 99949

QY 1479 CAACTCTGGGAAAGGCTTTGTAATTGGTGATAAAAAATTTGAATTTCTTGCATTTTCAT 1538
 DB 99950 CCATTCTAAAAAAGATTTTCCATTGGTCCCAAGACATTTGAATTTTGGCTTCTCAG 100009
 QY 1539 CGAGCCAGTTGCGGATATAATTCAGTGTGGATGTTTGCATCAAGACCTGSCCTTACTGCAA 1598
 DB 100010 CAACTCAACTCGGGGAAATCTCTTTGGATGTTTGGCTTATATGCTTCTTGAATGCTG 100069
 QY 1599 ATCATATAAGAGCTTGGATGGGTGATTTTTCGAGATCAAGAATGTCGCAAAATATGCTG 1658
 DB 100070 GAGTATAGAAGATGGATGGGCCACTTTCAGAAATATCCGTTCAAGTTCTAAATGTGCAG 100129
 QY 1659 CCAGACTTGGTCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTAGCATGAGA 1718
 DB 100130 CTAGAATGGGTCAAGCTTTCAGCTCTCCAGGCAACATTTGAAGTCTTACAAATGGATG 100189
 QY 1719 TTCAAGTTATTTCCGATGTAAGGT-----TCATGGAACCAAGCTATGTTCTTCTCGATG 1772
 DB 100190 TGAAGTATTCAGATATAGATCACACATGATGGTCCCAATACATATTTCTGATG 100249
 QY 1773 GAATGGTAAATATCTGGTGAATTTGCTCATAGATTTGCTCAAAATGTCGCTTCAAT 1832
 DB 100250 GTATTGGGAAGATATCTTTGAGATTTGCCAAGCGAGTTGCACACCACGTTGGATTAGATC 100309
 QY 1833 ATAC-----CCCATCTGCTTCCAGATTCGTTATGTTGATATAAGAGTGTGTGGGTG 1886
 DB 100310 CTACTAACCTTCTTCCAGCTTTCAAAATAGGTATGGTGGTACAAAGAGTCATTTGCCA 100369
 QY 1887 TTGATCCGGATTCATCAATGAAGTTGTCTTGAAGAAGAGCATGTCGAATATGAATCAG 1946
 DB 100370 TTGATCCTATGCTCTATTTGATCTTCTGCTGCTCTAGTATGAAGAAATTTGAGTCTG 100429
 QY 1947 ACAACATAAAGTTAGATGCTTGGATGGAGCAAAATATCAGCCTTGTATCTTAATTCGTC 2006
 DB 100430 AGAGCAGATGTTGAACATTTACAAGTTGGAGCAAGTCTCAGCCATGCTATGTAATCGTG 100489
 QY 2007 AACTGATTAGCTCTTGTCTACACTTGGAGTGAAGATGAAGTCTCGAACAGAGCAAA 2066
 DB 100490 AAATTTATTTCTCTCTCAACATTTGGTATAAGGATGAATTTTGTGCAATGCAAC 100549
 QY 2067 AGAAGCTGTAGATCAGTTGATGCTATCTTGATGATCTTTTGAAGGCACAGAGCTT 2126
 DB 100550 AAGATGAGATGCGTGAACAGAGAA---TGTACCAACAAAGAGTGTCTTTGTCAG 100606
 QY 2127 TGAATTTGATGCTCTCTGGAGAGAACACTAATTTCTCAGGCAATGCTAAATCTGGTT 2186
 DB 100607 TCCTAGGAAACTTGGTGGTTCTGAAACAAAGACAGCAGTGAAGATGTTCTGCAAGGT 100656
 QY 2187 ATAGCCTGATGCTGAGCCCTTTCTTCAATGATGTTGCAACCTTCCGCGCATCCAAGT 2246
 DB 100667 ATGAGCCTAGTTTCAGAGCCTTACCTGTCAATGATTTTAAAGGCCCATCAGGAAATAGG 100726
 QY 2247 TGCTCGATTTGCGGACTTAGATCAAGAAATATTTTCCAAATGGAAGAACAAATGATGGAT 2306
 DB 100727 TTACTGACATTAGAATAGATGTAATACATGTTCCAAAGGCCGTTGTTCTTATTGGT 100786
 QY 2307 GTTTGGATGAATCCAGAACCTTGGATATGGTCAGTGTGTTTGTTCAGTTTACTGTTGCTG 2366
 DB 100787 GTTTGGATGAACAGSTGTATTAGAATATGTCAGGTTTACATTCGAATTACAAAGAAC 100846
 QY 2367 GACATGAGAGTTTCTGACGATTTACATCCATTTTAAATACACAGATCCACCACAGTA 2426
 DB 100847 GCAAGAAC-----AAAGACAGTAATCAATCATATTTTATATGATGACGGGAAA 100900
 QY 2427 ATTTCATTTGAAGGAAATGTGTTGTTGCAAAAAATCCATGCTTCATCTCTGGTGATA 2486
 DB 100901 CAGCCACAGTTGTTGGAAGAGTTGCAATCAAAAAATCCGTTCTCCATCTCTGGTGATA 100960
 QY 2487 TTGCTGTTTAAAGCTGTAAATGTTGCGAGCTGCACACATGGTAGATTGTTGTTGAT 2546
 DB 100961 TCCGAGTACTTGAAGCCATCTATGACCTGATTGTTGGGCAAGTTGATTGCTTGTCT 101020
 QY 2547 TCCTCAGAAAGGAAAAAG 2563

```
||||| 101021 TCCCTCAAGAGGAGAAAG 101039
Db 101021 TCCCTCAAGAGGAGAAAG 101039

RESULT 13
AX204840
LOCUS AX204840 3591 bp DNA Linear PAT 30-AUG-2001
DEFINITION Sequence 2 from Patent WO0155407.
ACCESSION AX204840
VERSION AX204840.1 GI:15394184
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 3591)
Beclin,C., Elmayan,T., Mourrain,P. and Vaucheret,H.
Novel sgsl plant gene and use thereof
Patent: WO 0155407-A 2 02-AUG-2001;
Rhubio (FR) : INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (FR)
FEATURES
source
1..3591
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
1..3591
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC60174.1"
/db_xref="GI:15394185"
/translation="MGSEGNMKSVTVQVSGFGESTTAKQLDYLEDEIVWRCR
LKTSWTPPGSYNPFADSNIPSDIDYKKVPHAFVFAVESAGRAMDAAGCNLI
LDQPLKVSIGPKNPYSLNQRRITVYKLAGITLIGLTVLSRDFVSWRAEGVDFL
VDFDNTCKFCRKFSAFKDAVMHVNCDYKELLVRDIQTQVQYKTYLHGVLIL
QLASSPRVWRTADDIDYDTPGDLDDDDPWIRTDFTQVGAIGRCHSYRVLISPRY
ENKLRALDYFRMRVQERVWPPIRNPECGEPVSDHFCIHKEGISLEIMELV
SNVLHGVSNQFQTLERFDLLRNQKFNQNIASLKLKLYKRPVFDAYKRLKIQENI
QNKPLGSHQEODESEIRLRLVITRAYCLUPPEPELNSRLVRLKAVARFLRVTF
MDSMTIINSVLVFAIVKDLTSSFSQKTYFVKRYSILTDGFKLCGRKYSFLA
FSANQLRDSAMFADGKT RVSDIKTMGKFNKAKVAKARMGLCFSTATYDVM
PHEVDYEDIERNGVFSDIGITIPDLADEVWEKLDVHYSPQAIQIRVAGKPV
VAVPWSKDIQIRALRDSMKFKSKTILEICSTRFQPGFLNRQITLLSLVGPDE
IFWDMQESMLKLNRLDIDTDAFVLTASCAEQGNATAIMLSAGPKPTEPHRLGML
SVRIQALWGLRKSRIFFVTSGRWLMGLDEAGILEHGOCFTQVSPSENCESKHS
RFRETIDLEVVAGYVAIAKNPCLHPGDFVLEDAVDPQLHMLDCLIFPQGRPHI
NEASGLDGLDVFVADWOKLIPPNKSYPMHYDAEESKLSGRVNHQDIDIFARN
LANEQGLTICNAHVADRSQYGNKSYPMHYDAEESKLSGRVNHQDIDIFARN
LYDFMGKEDYQYKKNKILGRLYRRVKEVYDEDAEESSESTDPGAIYDAVLEIPG
FEDLIPWAGHKLYDGLIGLLGQYKVKOEIEIVTGHWSMPKYSKKQELKERLK
HSYNSLKKERKVFETIPDHENLSEENKLYEKKASWYHYVHPVKKSLLEQD
PDSSSHAMLSFANWIAADYLRKIKIRSENGSIDSAPVDSLAKFLAQL"
BASE COUNT 1022 a 714 c 856 g 997 t
ORIGIN

Query Match 8.2%; Score 305.8; DB 6; Length 3591;
Best Local Similarity 51.0%; Pred. No. 7e-62;
Matches 1013; Conservative 0; Mismatches 917; Indels 57; Gaps 10;

Qy 1263 TAAGAAGGTCCTAGTACACACCTGCAAGTTTATTTTGGTGGCCAGAGTTAAGTTT 1322
Db 1232 TCAGAAGACTAGTAATACCCACACAGAGCCTATTGCTTACCCACAGAGTTGAGTCT 1291
Qy 1323 CCAATCGGTTCTCGGGAATATTCTGAAGACATAGATACTTTCTTCTGTTCTTTT 1382
Db 1292 CCACAGGGTACTCAGGAGATACAAAGCTGTGCTGAAAGATTTTTCGGGGTAACTTCA 1351
Qy 1383 TTGATGAGGAGTGGGAGAACTGTATTC-----TACAGACTTATTACCAGAA 1429
Db 1352 TGGATGAAAGTATGACAGACCAATAATTCGAATGTTCTCTTACTTCTTCTTCTT 1411
Qy 1430 GCAAGTACTGGAAGTGGT-----CAGGACAAACATCTATGAGAGGATCTTAT 1478
Db 1412 TGAAGGATTCGACATCAAGTTCCTTCTCCAGAGACCTAGCTTTTAAAGAGTGAAGA 1471
```

```
QY 1479 CAACCTCTCGGGAAGGCTTTGTAATTTGGTGTGATAAAAAATTTGAATTTCTTGTGATTTTCAT 1538
Db 1472 GCATATTAACCGATGGGTTTAAACTATGTGGTAGAAAAACACAGTTCTTCTAGCATTTCTCAG 1531
QY 1539 CGAGCCAGTTGGGGGATAAATTCAGTGTGGATGTTTGTCAAGACCTGGCCCTTACTGTCAA 1598
Db 1532 CCNATCACTGAGAGACCCCTCTGCATGGTCTTTGCTGAAGACGGGAAAAACACGTGTGT 1591
QY 1599 ATGATATAAGAGCTTGGATGGGTGATTTTCGAGATCAAGAAATGTCGCAAAATATGCTG 1658
Db 1592 CAGATATAAAAAACATGGATGGGGAAGTTCAAGA---CAAGAATGTGCAAAATATGCTG 1648
QY 1659 CCAGACTTGGTCAACTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTATAGCATGAGA 1718
Db 1649 CTAGGATGGGCTGTGCTTCTCTCCACATATGCCACTGTAGATGTCATGCCCTCAGGAG 1708
QY 1719 TTGAAGTTATTCGCCGATTAAGGTTTCATGGAACACAGCTATGCTTTTCTGTATGGAATG 1778
Db 1709 TTGA---CACTGAGGTTCCAGATATTGAGAGAAATGGGTATGTTTCTCTGACGGAATG 1765
QY 1779 GTAAAAATATCTGTGACTTTGCTCATAGAGTTGCCCTCAAAA-----TGTGGCCTTCAAT 1832
Db 1766 GTACAATCACACTGACCTCGCTGAGCAAGTAAATGAGAAACTTAAGTTGGATGTGCACT 1825
QY 1833 ATACCCCATCTGCTTCCAGATTCGTTATGTTGGATATAAAGGIGTGTGGGTGTTGATC 1892
Db 1826 ACAGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAGGGGTTGTTGCTCGTTGC 1885
QY 1893 CGGATTCATC-----AATGAAGTTGCTTTTGAAGAGAGCATGCGAATATGAATCAG 1946
Db 1886 CATCAAAAAGTATGAGAAATCAGCTAGCCCTTCGAGACAGTATGAAGAAAGTCTTTTCCA 1945
QY 1947 ACAACATAAAGTTAGATGCTCTTGGATGGAGCAAAATATCAGCCTTGTATCTTAATCGTC 2006
Db 1946 AACATACGATCTTGGAGATCTGTTCTCGAGAGGTTTCAACCTGGGTTCTTAAATCGGC 2005
QY 2007 AACTGATTAGCTCTTGTCTACACTTGGAGTGAAGATGAAGTTCTCGACACAGAACAA 2066
Db 2006 AGATAATACCCCTTCATCCGACTAGGTGTTCCGGGATGAATATTTCTGGGATATCAGG 2065
QY 2067 AGGAAGCTGTAGATCAGCTTGTATCTTTCATGATGTTGCAAAACCTTCCGCGCATCCAAGT 2246
Db 2066 AATCCATGCTCTATAAATGAACCGCATCTTGTATGATACAGATGTGGCATTTGAAGTTC 2125
QY 2127 TGAATTTGATGCTCTCTGGAGAGAACACTAATATTTCTCAAGCAATGCTAAACTGTGTT 2186
Db 2126 T---CACGGCATCATGCTGTGAACAGGGAACACTCGACTATCATGCTTAGTGCAGGTT 2182
QY 2187 ATAAGCTGTATGCTGAGCCCTTTCTTTTCAATGATGTTGCAAAACCTTCCGCGCATCCAAGT 2246
Db 2183 TCAACCAAAAACCGAGCCGCATCTACGGGGATGTTGCTTTCAGTCAGATTTGCACAAC 2242
QY 2247 TGCTCATTTGGGACTAGATCAAGAAATATTATTTCCAAATGGAAGAACAAATGATGGAT 2306
Db 2243 TCTGGGCTCTCAGAGAAAAATCTCGTATTTTGTACTTCAGGAAGGTGGCTAATGGGT 2302
QY 2307 GTTTGGATGAATCCAGAACCTTGGAAATATGTTGAGGTGTTTGTTCAGTTTACTTGTGCTG 2366
Db 2303 GCTTACAGAGAGCGGATACITTGACATGGCCAAATGCTTTATTCAAGTCTCTAAACC-- 2360
QY 2367 GACATGAGAGTTTCTGACGATTTACATCCATTTAATACAGCAGATCCACCACAGTA 2426
Db 2361 -GTCTATAGAAAATTTGTTCTCCAAAACATGGGTCTCGTTTTTAAAGAGACAAAGACAGATC 2419
QY 2427 ATTTCAATTCGAGGGAATGTTGTTGCCAAAAATCCATCTTGCATCTCTGTTGATA 2486
Db 2420 TGAAGTAGTTAAGGCTATGTAGCCATTTGTCAGCAATCTTGTCTTCCACAGGGGATG 2479
QY 2487 TTCGTGTTTAAAGGCTGTAAATTTTCGAGCGCTGCACACATGTTGATGTTGTTGAT 2546
Db 2480 TAAGGATTTTGAAGCTGTTGATGTACCCAGCTGCATCAGATGATGACTGCTTATTT 2539
```


QY	1736	GTAAAGTTTCATGGAAACAGCATGCTCTTTCTGATGGAATGGTAAATATCTGGTGAC	1795
DB	112231	GATGTGTGCAT---AATGAGTACATTTCTCCGATGAATTTGGCAAAATATCCACACAT	112175
QY	1796	TTTGCTCATAGATGGCTCAAAATGTGGCTTCATATAC--CCATCTGCTTTCCAG	1852
DB	112174	CTTGCAATGGAAAGTTGCCGAGAGGCTGCAGCTTGACAGATAACCCCATCTGCTTATCAG	112115
QY	1853	ATTCGTTATGGTGATATAAAGTGTGTGGGTG-----TGATCCGATTCATCAATG	1906
DB	112114	ATCAGGTTTGGCTTCAAGGTTGTCATAGCTGTCTGCAAGGACATGGTGATGGGACA	112055
QY	1907	AGTTGCTTTTGGAGAAAGATGTCGAAATATGATCAGACAAATAGTATGATGTC	1966
DB	112054	CGGCTTTTCTGAGCCAAAGATGAGGAGTTGAGCTTAACCATTTGGTTTAGAGTG	111995
QY	1967	CTTGATGAGCAAAATATCAGGCTTGTATCTTAATCTGCTCACTGATACGCTTGCT	2026
DB	111994	GTCCTCGACAAAGTCCAGCCAGGATCTTAAATCGACATATAATATTGCTATCC	111935
QY	2027	ACATCTGGAGTGAAGATGAAGTTCTCGAACAGACAAAGAGCTGTAGATCAGTT	2086
DB	111934	TCAGTGAATGTCCAGATCTTATCTTTTGGCAATGCAAGACCATGCTTTCTAACCTC	111875
QY	2087	GATGCTATCTTGATGATCTTTGAAGGCACAGAGGCTTTTGAATTTGATGCTCTCGA	2146
DB	111874	AACAATATCTATCAGACAGAGATGTTGCTTTTGAAGTTTAAACA---ACTTCATGTGT	111818
QY	2147	GAGAACACTAATATCTCAAGCAATGCTAACTGTGTTTAAAGCTGATGCTGAGCCC	2206
DB	111817	GATGATGGAAACACTCAGCATGATGCTCAGTGTGGCTTTGAACCTAGATGAACCA	111758
QY	2207	TTCTTTCAATGATGTTGCAAACTTCCCGCATCCAGATTTGCTGATTTGGGGACTAGA	2266
DB	111757	CATTGAAACAATGCTCTTGGCTAATAGTCTGACAAATTCAGGATCTTTTGGAAAA	111698
QY	2267	TCAAGAAATTTATCCAAATGGAAGAACATGATGGGATGTTTGGATGAATCCAGAAC	2326
DB	111697	GCAAGGATATTTGTGCAAGGAAGGTGTTGATGGCTGCTTTGATGAGCTTGGGPT	111638
QY	2327	TTGAAATATGTCAGGTGTTTGTAGTTTACTGTTGCTGGATGGACATGGAGTTTCTGAC	2386
DB	111637	CTTGAGCAAGGAGTGCTTTATTCGGGCAACAGTCCATCATGATAGTATTTTGT	111578
QY	2387	GATTACATCCATTTAATACACAGATCCACACAGTAATTTCAATCTGAAGGAAAT	2446
DB	111577	AAGCATGGGTCAAGATTTTCATCAACAGATATAAACACACAGAGTCAATTT---GGTACT	111521
QY	2447	GTGTTGTTTCCAAAATCATCTTCCATCTGCTGGTATATGCTTTTAAAGGCTGTA	2506
DB	111520	GTGTAATGCAAGAAATCCCTGCTTCATCCAGGGATGTCGCAATCCTTGAGCAGTT	111461
QY	2507	AATGTTCCGAGGCTGCACCATGAGTGTGTTGATTCCTCCAGAAAGAAAGA	2566
DB	111460	GATGTGCCGAACATGCATCATCTGTTGATTTGTTGTTCCCCAGAAAGGTGAGAG	111401
QY	2567	CTCATCCGAATGAATGTTCTGGAGTATTTGGATGGGATATCTACTTTGTTGCTGG	2626
DB	111400	CCATGCTAACAGGAGTCTGGAGGATCTTTGATGGGATCTCTACTTTGTGACATGG	111341
QY	2627	GATCAAGACATGATCCGCCCAAG	2649
DB	111340	GATGAGAACTTATACCTCCAGG	111318

RESULT 15
AF239718
LOCUS AF239718 4013 bp DNA linear PLN 02-JUN-2000
DEFINITION Arabidopsis thaliana SGS2 gene, complete cds.
ACCESSION AF239718
VERSION AF239718.1 GI:8164027
KEYWORDS Arabidopsis thaliana.

ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 4013)			
AUTHORS	Mourrain,P., Beclin,C., Elmayan,T., Feuerbach,F., Godon,C., Morel,J.B., Joutette,D., Lacombe,A.M., Nikic,S., Picault,N., Remoue,K., Sanial,M., Vo,T.A. and Vaucheret,H.			
TITLE	Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance			
JOURNAL	Cell 101 (5), 533-542 (2000)			
MEDLINE	20306668			
PUBMED	10850495			
REFERENCE	2 (bases 1 to 4013)			
AUTHORS	Beclin,C., Mourrain,P., Vaucheret,H. and Elmayan,T.			
TITLE	Direct Submission			
JOURNAL	Submitted (28-FEB-2000) Biologie Cellulaire, INRA, Route de Saint-Cyr, Versailles 78026, France			
FEATURES	Location/Qualifiers			
source	1. 4013			
organism	"Arabidopsis thaliana"			
cultivar	"Columbia"			
db_xref	"taxon:3702"			
chromosome	"III"			
join	join(<1.. 2715,3138.. >4013)			
product	/product="SGS2"			
join	join(1.. 2715,3138.. 4013)			
note	/note="RNA-dependent RNA polymerase-like protein"			
codon_start	1			
product	/product="SGS2"			
protein_id	/protein_id="AA073959.1"			
db_xref	/db_xref="GI:8164028"			
translation	"MGSEGNKKSVTVQVSGFGESTTAKQLTDVLEDEVGIVWRCR LKTSWTPGSPNPFADTNIPTIDYKVEHPHFVAFVAFESAGRAMDAAGCCNL LDQGLKVSGLGNPNYSLNORRTTTPYKLAGITLTVGTLVSRDDDFVSWRAGVDFL VDFDNCCKFKSTAFKDAVMAHVIKLELLVDRDIQTQYKTLHGFVLIL QLAASSPRVYRTADDIDYTPVGLDDDDPWIRTDFTQVGAIGCHSYRLISPRY ENKRLTALDYFRMRVQEVRRPPIRNPCEPGEVSDHFFCIHKHKGISFFIMFLV NSVLRGVFNQQLTERFPELRLNPKDVAISLHCTVRYKRVKVAERFLRVTF QKNPKLGLSHQESDISEIRRLVITPRAYCLPPEVLSNRLLRVRKVAERFLRVTF MDESMQITNSNVLVSFVAPIVKLTSSFSQKTVYFKRVKSILTDGFKLGRKYSFLA FSNQLRDSAGWFAEDGKTRVSDITMWSKFKDKNVAKCAARMGUCFSTYATVDM PHEVDTVPDIERNYVYFSDIGITITPDLADEYMEKLLKDVHYSPOAYQIRYAFKGV VAWPSSDGIIRALRDSMKKFFSKHTILETCSWTFQPLNRQITTLUSVLGVPDE IFWDMOSMLYKLNRIILDDVAFVLVATSCAEOGTAAIMLSAGFKPTEPHLGM LSVRIAQIWLGRKSRIFVTPSGWLGCIDEGILEHGOCFIOVSKPSIENCSKHGS RFKRTKTDLEVVQIVAIKANKPLHPDGVRLIENVDPQLHHVDCILFIPQKGRPHIT NEASGSLDGLDYEVWDQKLIPPNRKSYPAMHYDAAEKSLGRVANHDIIDFARN LANQLGTICNAHYVHADRSYEGAMDEECILLAEATAVDFPATGIYVSNPHLKPK LYPDFMGKEDYQTYKSKILGRLYRVYKVEDAEASEEDSPSAIPDAVLEIFG FEDLIPEAWGHKCLYDGLIGLQYKQVEEIVTGHIMSPRYTSKKOGELKERUK HSYNSLKKERKVFETIPDHENISEEKNILYKASAWHVTYHPEWYKSLQLOD PDESSHAMLSFAMTAADYLARIKIRSRMGSIDSKRPVDSLAKLAQL"			
BASE COUNT	1119 a	778 c	948 g	1166 t
ORIGIN				
Query Match	5.9%;	Score 220.4;	DB 8;	Length 4013;
Best Local Similarity	51.5%;	Pred. No. 1.4e-41;		
Matches 733; Conservative	0;	Mismatches 641;	Indels 48;	Gaps 8;
QY	1263	TAAGAAGGTCCTAGTACACCATGCAAGTTTATTTTGTGTCGACAGGTTAATGTTT	1322	
DB	1232	TCAGAAGACTAGTAATACCCCAACAGAGCCTATTCCTGCCATCCCAAGAGTGTGCT	1291	
QY	1323	CCAATCGGGTTCCTCGCAATTTCTGAGACATAGATAAACTTTCTCGTGTCTTTTG	1382	
DB	1292	CCAACAGGGTACTCAGGAGATACAAAGCTGTTGCTGAAAGATTTTTCGGGTAACCTCA	1351	
QY	1383	TTGATGAGAGTGGGAGAACTGTATTC-----TACAGACTTATACAAAA	1429	
DB	1352	TGGATGAAAGTATGCAGACCAATAAATTCGAATGTTCTCTTACTTTGTGCTCTATTG	1411	
QY	1430	GCAAGTACTGGAAGTGGTGT-----CAGGACAAACATCTATGAGAGATCTTAT	1478	

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 4013)

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C.,
Morel, J. B., Joutette, D., Lacombe, A. M., Nikic, S., Picault, N.,
Remoue, K., Sanial, M., Vo, T. A. and Vaucheret, H.

Arabidopsis SGS2 and SGS3 genes are required for
posttranscriptional gene silencing and natural virus resistance
Cell 101 (5), 533-542 (2000)

20306668
10850495
2 (bases 1 to 4013)

Beclin, C., Mourrain, P., Vaucheret, H. and Elmayan, T.
Direct Submision
Submitted (28-FEB-2000) Biologie Cellulaire, INRA, Route de
Saint-Cyr, Versailles 78026, France

Location/Qualifiers
1. 4013
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="III"
join(1.. 2715,3138.. >4013)
/product="SGS2"
join(1.. 2715,3138.. 4013)
/note="RNA-dependent RNA polymerase-like protein"
/codon_start=1
/product="SGS2"
/protein_id="AA073959.1"
/db_xref="GI:8164028"

translation="MGSEGNKKSVTVQVSGFGESTTAKOLTDVLEDRVIVWRCR
LKTSWTPGSPNPFADTNIPTIDYKVEHPHFVAFVAFESAGRAMDAAGCCNL
LDQGLKVSGLGNPNYSLNORRTTTPYKLAGITLTVGTLVSRDDDFVSWRAGVDFL
VDFDNCCKFKSTAFKDAVMAHVIKLELLVDRDIQTQYKTLHGFVLIL
QLAASSPRVYRTADDIDYTPVGLDDDDPWIRTDFTQVGAIGCHSYRLISPRY
ENKRLTALDYFRMRVQEVRRPPIRNPCEPGEVSDHFFCIHKHKGISFFIMFLV
NSVLRGVFNQQLTERFPELRLNPKDVAISLHCTVRYKRVKVAERFLRVTF
QKNPKLGLSHQESDISEIRRLVITPRAYCLPPEVLSNRLLRVRKVAERFLRVTF
MDESMQITNSNVLVSFVAPIVKLTSSFSQKTVYFKRVKSILTDGFKLGRVSLA
FSNOLDRSAGWFAEDGKTRVSDITMWSKFKDKNVAKCAARMGUCFSTYATVDM
PHEVDTVPDIERNYVYFSDIGITITPDLADEYMEKLLKDVHYSPOAYQIRYAFKGV
VAWPSSDGIIRALRDSMKKFFSKHTILETCSWTFQPLNRQITTLUSVLGVPDE
IFWDMOSMLYKLNRIILDDVAFVLVATSCAEOGTAAIMLSAGFKPTEPHLGM
LSVRIAQIWLGRKSRIFVTPSGWLGCIDEGILEHGOCFIOVSKPSIENCSKHGS
RFKRTKTDLEVVQIVAIKANKPLHPDGVRLIENVDPOLHHVDCILFIPQKGRPHIT
NEASGSLDGLDYEVWDQKLIPPNRKSYPAMHYDAAEKSLGRVANHDIIDFARN
LANQLGTICNAHYVHADRSYEGAMDEECILLAEATAVDFPATGIYVSNPHLKPK
LYPDFMGKEDYQTYKSKILGRLYRVYKVEDAEASEEDSPSAIPDAVLEIFG
FEDLIPEAWGHKCLYDGLIGLQYKQVEEIVTGHIMSPRYTSKKOGELKERUK
HSYNSLKKERKVFETIPDHENISEEKNILYKASAWHVTYHPEWYKSLQLOD
PDESSHAMLSFAMTAADYLARIKIRSRMGSIDSKRPVDSLAKLAQL"

BASE COUNT 1119 a 778 c 948 g 1166 t

ORIGIN

Query Match 5.9%; Score 220.4; DB 8; Length 4013;
Best Local Similarity 51.5%; Pred. No. 1.4e-41;
Matches 733; Conservative 0; Mismatches 641; Indels 48; Gaps 8;

QY 1263 TAAGAAGGTCCTAGTACACCATGCAAGTTTATTTTGTGTCGACAGGTTAATGTTT 1322
DB 1232 TCAGAAGACTAGTAATACCCCAACAGAGCCTATTCCTGCCATCCCAAGAGTGTGCT 1291

QY 1323 CCAATCGGGTTCCTCGCAATTTCTGAGACATAGATAACTTTCTCGTGTCTTTTG 1382
DB 1292 CCAACAGGGTACTCAGGAGATACAAAGCTGTTGCTGAAAGATTTTTCGGGTAACCTCA 1351

QY 1383 TTGATGAGAGTGGGAGAACTGTATTC-----TACAGACTTATACAAAA 1429
DB 1352 TGGATGAAAGTATGCAGACCAATAAATTCGAATGTTCTCTCTTACTTTGTGCTCTATTG 1411

QY 1430 GCAAGTACTGGAAGTGGTGT-----CAGGACAAACATCTATGAGAGATCTTAT 1478

Db 1412 TGAGGATTTGACATCAAGTCTCTTCCAGAGACCTACGTTTAAAGAGTGAAGA 1471
Qy 1479 CAACCTCGCGAAGAGGCTTTGTAATGGTGATATAAAATTTGAATTTCTTGGCAATTTTCAT 1538
Db 1472 GCATTTAAACCGATGGGTTTAAACTATGTGTAGAAAATACAGTTTCTAGCAATCTCAG 1531
Qy 1539 CGAGCCAGTTGCGGGATTAATTCAGTGGGATGTTTCATCAAGACCTGGCCCTTACTGCAG 1598
Db 1532 CCAATCAACTGAGAGACCGCTCTGCATGGTCTTTGCTGAAGACGGGAAAACACGTTGT 1591
Qy 1599 ATGATATAAGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAAATGTGCGCAAAATATGCTG 1658
Db 1592 CAGATATAAACAATGATGGGAAGTT---CAAAGACAAGAATGTGGCAAAATGTGCTG 1648
Qy 1659 CCAGACTTGGTCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTTAGSCATGAGA 1718
Db 1649 CTAGATGGGCGCTGTCTCTCCACATATGCCACTAGATGTCATGCCCTCACGAGG 1708
Qy 1719 TTGAAGTTATTCGGATGTAAGGTTTCATGGAACCACTATGTCTTTTCGATGGAAATG 1778
Db 1709 TTGA---CACTGAGTTCCAGATATTGAGAGAAATGGGTATGTTTCTCTGACGGAATG 1765
Qy 1779 GTAAAAATATCTGGTGAATTTGCTCATAGAGTTGGCTCAAAA-----TGTGGCCCTCAAT 1832
Db 1766 GTACAATCACACCTGACCTCGCTGACGAAGTAAATGGAGAAACTTAAGTTGATGTGCACT 1825
Qy 1833 ATACCCCATCTGCTTTCAGATTCGTTATGTTGATATAAGGTTGTTGSGTGTTCATC 1892
Db 1826 ACAGCCCTGTGCTTATCAGATAGCTTACGAGGTTTCAAGGGGTTGTTGCTCGTGGC 1885
Qy 1893 C-----GGATTCATCAATGAAGTTGCTTTTGAAGAGAGCATGCGAAATATGAATCAG 1946
Db 1886 CATCAAAAAGTGAATCAAGCTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCA 1945
Qy 1947 ACAACATTAAGTTAGATGCTTTGGATGGAGCAATATCAGCCCTGTATCTTAATCGTC 2006
Db 1946 AACATACGATCTTGAGATCTGTTCCCTGGAGAGGTTTCAACCTGGGTTCTTAAATCGGC 2005
Qy 2007 AACTGATTACGCTCTTGTCTACACTTGGAGTGAAGATGAAGTTCTCGAAGCAAGCAAA 2066
Db 2006 AGATTAATTAACCTTCTATCCCTACTAGTGTCCGGATGAATATCTGGGATATGCAGG 2065
Qy 2067 AGAAGCTGTAGATCAGCTTATGCTATCTTGCATGATCTTTTGAAGGCACAGGAGCTT 2126
Db 2066 AATCCATGCTCTATAAATCAACCGCATCTTGTATGATACAGATGTGSCATTTGAAGTTC 2125
Qy 2127 TGGATTGATGTCCTCGAGAGAACACTAATATTTCAAGGCAATGCTAAACTGTGGTT 2186
Db 2126 TCACGGCATCATGTCTGAACAGGAAACACTGCAGCTAT---CATGCTTAGTCAGGTT 2182
Qy 2187 ATAGCCCTGATGCTGAGCCCTTTCTTCAATGATGTGTCAAACCTTCGCGCATCCCAAGT 2246
Db 2183 TCAAAACCAAAAACCGAGCCGATCTACCGGGATGTTGCTTCACTCAGAAATGCACAAC 2242
Qy 2247 TGCTCGATTTGGGACTAGATCAGAAATATTATCCAAATGGAGAACAAATGATGGAT 2306
Db 2243 TCTGGGGTCTCAGAAAAATCTCGTATTTTGTACTTTCAGGAGGTTGGCTAATGGGTT 2302
Qy 2307 GTTTGGATGAATCCAGAACCTTGAATATGTCAGTGTGTTTGTTCAGTTTACTGTGCTG 2366
Db 2303 GCCTAGACGACGAGGATACTTGAACATGGCAATGCTTTATTCAGTCTCTAAACCGT 2362
Qy 2367 GACATGGAGAGTTTCTGACGATTTACATCCATTTAATAACAGCAGATCCACCACAGTA 2426
Db 2363 CTATAGAAAATTTGTTCTCCAAACATGGGTCTCGTTTTTAAGAGACAAAGACAGATCTGG 2422
Qy 2427 ATTTCAATCTGAAGGAAATGTTGTTGCAAAAAATCCATGCTTGCATCTGCTGTGATA 2486
Db 2423 AAGTAGT---AAGGCTATGTAGCCATGCTAAGATCCCTGTCTTCACCCAGGGATG 2479
Qy 2487 TTCGTGTTTAAAGCTGTAAATGTTCGAGCCCTGCACACATGCTAGATGTGTTGAT 2546

Db 2480 TAAGGATTTTAGAAGCTGTTGATGTACCCAGCTGCATCACATGTATGACTGCCCTTATTT 2539
Qy 2547 TCCTTCAGNAAGAAAAGACCTCATCCGAATGAATGTTCTGGGAGTGTATTGGATGGGG 2606
Db 2540 TCCTTCAGAAAGGTATAGGCCCGCATACAAAGCAAGCTTCTGCGAGTGACCTTGACGGG 2599
Qy 2607 ATATCTACTTTTGTGTTGCTGGGATCAAGACATGATCCCGCCAA 2648
Db 2600 ACCTGTACTTTTGTGGCTTGGGATCAGAAACTCATCCCTCCCA 2641

Search completed: November 5, 2002, 22:34:30
Job time : 11044 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: November 5, 2002, 19:16:42 ; Search time 142 Seconds
(without alignments)
8057.824 Million cell updates/sec

Title: US-09-782-874-1
Perfect score: 3731
Sequence: 1 GAATATCTTCTTACTT.....AGTTTCATCTTCTTCTTAA 3731

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3731	100.0	3731	4 US-08-811-583-1	Sequence 1, Appl
2	57.2	1.5	906	1 US-08-100-874-1	Sequence 1, Appl
3	53.4	1.4	7218	1 US-08-232-463-14	Sequence 14, Appl
4	40	1.1	162450	1 US-09-345-882-1	Sequence 1, Appl
5	36.8	1.0	2277	1 US-08-676-967-2	Sequence 2, Appl
6	36.8	1.0	2277	1 US-08-676-974-2	Sequence 2, Appl
7	36.8	1.0	2277	2 US-09-098-487-2	Sequence 2, Appl
8	36.6	1.0	7218	1 US-08-232-463-14	Sequence 14, Appl
9	36.4	1.0	3191	4 US-08-936-165A-194	Sequence 194, Appl
10	35.4	0.9	17341	4 US-09-415-946-1	Sequence 1, Appl
11	35	0.9	1248	4 US-09-134-001C-342	Sequence 342, Appl
12	35	0.9	1275	2 US-08-911-445-4	Sequence 4, Appl
13	35	0.9	1275	4 US-09-182-983-4	Sequence 4, Appl
14	35	0.9	1458	2 US-08-911-445-19	Sequence 19, Appl
15	35	0.9	1458	4 US-09-182-983-19	Sequence 19, Appl
16	35	0.9	1796	2 US-08-911-445-17	Sequence 17, Appl
17	35	0.9	1796	4 US-08-182-983-17	Sequence 17, Appl
18	35	0.9	1845	2 US-08-911-445-5	Sequence 5, Appl
19	35	0.9	1845	4 US-09-182-983-5	Sequence 5, Appl
20	35	0.9	2046	2 US-08-911-445-6	Sequence 6, Appl
21	35	0.9	2046	4 US-09-182-983-6	Sequence 6, Appl
22	35	0.9	5613	2 US-08-463-418-1	Sequence 1, Appl
23	35	0.9	111282	4 US-09-754-250-3	Sequence 3, Appl
24	34.8	0.9	1028	4 US-08-858-207A-222	Sequence 222, Appl
25	34.8	0.9	6002	4 US-09-345-882-4	Sequence 4, Appl
26	34.8	0.9	19446	4 US-08-961-527-51	Sequence 51, Appl
27	34.6	0.9	1883	4 US-08-991-677-3	Sequence 3, Appl

28	34.6	0.9	19390	4 US-08-961-527-86	Sequence 86, Appl
29	34.4	0.9	738	4 US-09-004-838-62	Sequence 62, Appl
30	34.2	0.9	1003	4 US-09-221-017B-395	Sequence 395, Appl
31	34.2	0.9	2694	1 US-08-465-995A-1	Sequence 1, Appl
32	34.2	0.9	2694	2 US-08-465-994C-1	Sequence 1, Appl
33	34.2	0.9	2694	2 US-08-986-145-1	Sequence 1, Appl
34	34.2	0.9	2760	1 US-08-101-593-1	Sequence 1, Appl
35	33.8	0.9	258	4 US-09-345-882-21	Sequence 21, Appl
36	33.8	0.9	1867	4 US-09-377-557-3	Sequence 3, Appl
37	33.8	0.9	2889	4 US-09-134-001C-1660	Sequence 1660, Appl
38	33.6	0.9	2870	4 US-09-134-078-59	Sequence 59, Appl
39	33.6	0.9	3707	1 US-08-118-101A-1	Sequence 1, Appl
40	33.6	0.9	10223	4 US-08-961-527-73	Sequence 73, Appl
41	33.4	0.9	1939	1 US-08-232-188-3	Sequence 3, Appl
42	33.4	0.9	1939	3 US-08-970-723-3	Sequence 3, Appl
43	33.4	0.9	1939	5 PCT-US95-02708-3	Sequence 3, Appl
44	33.4	0.9	2026	3 US-08-755-587-26	Sequence 26, Appl
45	33.2	0.9	3636	4 US-09-134-001C-1983	Sequence 1983, Appl

ALIGNMENTS

RESULT 1
US-08-811-583-1
; Sequence 1, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811.583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3731 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORIGINAL SOURCE:
ORGANISM: Tomato
FEATURE:
NAME/KEY: CDS
LOCATION: 194..3535
US-08-811-583-1

Query Match 100.0%; Score 3731; DB 4; Length 3731;				
Best Local Similarity 100.0%; Pred. No. 0;				
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	GAATAATCTCTTACTTACTTACCCAGGAGTTGACTCATCAGTCCCTCAAGTCTTTTGT	60	
Db	1	GAATAATCTCTTACTTACTTACCCAGGAGTTGACTCATCAGTCCCTCAAGTCTTTGT	60	
Qy	61	GTGTGATATAAATTTGGTTGTCTTCAGTTTCAGTCTACTGCTGGTAGTTTAT	120	
Db	61	GTGTGATATAAATTTGGTTGTCTTCAGTTTCAGTCTACTGCTGGTAGTTTAT	120	
Qy	121	TTTGCAATAACTTCAGGGGGTATTCAGTTTGTAGCATTTTGAAGTGGAACTGCACAT	180	
Db	121	TTTGCAATAACTTCAGGGGGTATTCAGTTTGTAGCATTTTGAAGTGGAACTGCACAT	180	
Qy	181	GGATTTGGCTACATGGGAAGCAATTCAGTTTTCGGATTCCTTATCTCTCTGC	240	
Db	181	GGATTTGGCTACATGGGAAGCAATTCAGTTTTCGGATTCCTTATCTCTCTGC	240	
Qy	241	GGAGTGGTTAAGTCAATCTTAGAGAAATATACAGGATTTGGAACGTATGTGCATTTGA	300	
Db	241	GGAGTGGTTAAGTCAATCTTAGAGAAATATACAGGATTTGGAACGTATGTGCATTTGA	300	
Qy	301	GGTTAAACAGTCCAAAGGAGTCTAGACATTTGCCAAGTTCAATTTGCCGACACAT	360	
Db	301	GGTTAAACAGTCCAAAGGAGTCTAGACATTTGCCAAGTTCAATTTGCCGACACAT	360	
Qy	361	AAGTGCATCAAAATCATCACTTTGGCTAATAACAGCTGTATTTGGCTCTCTTATTT	420	
Db	361	AAGTGCATCAAAATCATCACTTTGGCTAATAACAGCTGTATTTGGCTCTCTTATTT	420	
Qy	421	GAAGGCTTGGAAATGAAACTGATATTGCCAACTGCCGGCATATGTGGATCAGATGA	480	
Db	421	GAAGGCTTGGAAATGAAACTGATATTGCCAACTGCCGGCATATGTGGATCAGATGA	480	
Qy	481	TGGCATCACTTTGAATTCGGATGTCAGATATCAGATGACAACTTTGCCAGTGTGGAG	540	
Db	481	TGGCATCACTTTGAATTCGGATGTCAGATATCAGATGACAACTTTGCCAGTGTGGAG	540	
Qy	541	TACAGAAGTTTCAATTCATTTGSCATTTGGATTGAAGAAATTTTCTTTTATCTAG	600	
Db	541	TACAGAAGTTTCAATTCATTTGSCATTTGGATTGAAGAAATTTTCTTTTATCTAG	600	
Qy	601	TGGTTACGTGACTATAAATTCAGCTTTTCATATGAATATATGACAGTTTGTCTCCA	660	
Db	601	TGGTTACGTGACTATAAATTCAGCTTTTCATATGAATATATGACAGTTTGTCTCCA	660	
Qy	661	TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTGGTGTCTCTCGAT	720	
Db	661	TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTGGTGTCTCTCGAT	720	
Qy	721	CTATAAGAGACTTGAAACCTCCTGTTATAGCTTCTTTAAGGAAACTCCTGATCAGTG	780	
Db	721	CTATAAGAGACTTGAAACCTCCTGTTATAGCTTCTTTAAGGAAACTCCTGATCAGTG	780	
Qy	781	GGTGAGACACAGATTTCCCTCATCTTGGATAGGCTATCTTCAGCTTATGTTTGA	840	
Db	781	GGTGAGACACAGATTTCCCTCATCTTGGATAGGCTATCTTCAGCTTATGTTTGA	840	
Qy	841	GTTCCTGTAGGGGTGTTCTCTTCCAAATTTTCAGGAAAGTTTTCACATATCCAGACG	900	
Db	841	GTTCCTGTAGGGGTGTTCTCTTCCAAATTTTCAGGAAAGTTTTCACATATCCAGACG	900	
Qy	901	TGAAAAAATTAATTTACTTACAGACTGGTTTCACCTTTTTCGTCTCTCAAAAATTCGGCTCT	960	
Db	901	TGAAAAAATTAATTTACTTACAGACTGGTTTTCACCTTTTTCGTCTCTCAAAAATTCGGCTCT	960	
Qy	961	GGTTCCCAATGTCCAGCTCCGGAGGAATTTCAATTCCTTACAGATTTTGTTCAAAAT	1020	
Db	961	GGTTCCCAATGTCCAGCTCCGGAGGAATTTCAATTCCTTACAGATTTTGTTCAAAAT	1020	
Qy	1021	TAGTCTTTTGGTACAGCATGGATGATACCTGGCCAGCATTAATGTCTACTTTTCCG	1080	

Db	1021	TAGTCTTTTGGTACAGCATGGATGATACCTGGCCAGCATTAATGTCTACTTTTCCG	1080	
Qy	1081	ATTAGTTGATCCTCGAAGAGAAATGTGGCATGCAATTGAGCATGCCCTTAGAAGAACTGA	1140	
Db	1081	ATTAGTTGATCCTCGAAGAGAAATGTGGCATGCAATTGAGCATGCCCTTAGAAGAACTGA	1140	
Qy	1141	CTATATAAAGAGTCTGTTATGATCCCGTGGGTGGCTCACTGAGCAGTATGATGGTA	1200	
Db	1141	CTATATAAAGAGTCTGTTATGATCCCGTGGGTGGCTCACTGAGCAGTATGATGGTA	1200	
Qy	1201	TCTCAAGGTTAGACAACCTCCAAAATCTCCGTCCTCACTTTAGATGGTGGTGA	1260	
Db	1201	TCTCAAGGTTAGACAACCTCCAAAATCTCCGTCCTCACTTTAGATGGTGGTGA	1260	
Qy	1261	TCTAAGAGGGTCTAGTAACACCATGCAAAAGTTTATTTTGTGTCAGAGGTTAATGT	1320	
Db	1261	TCTAAGAGGGTCTAGTAACACCATGCAAAAGTTTATTTTGTGTCAGAGGTTAATGT	1320	
Qy	1321	TTCCAAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTCGTGTCTTT	1380	
Db	1321	TTCCAAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTCGTGTCTTT	1380	
Qy	1381	TGTTGATGAGGAGTGGGAGAACTGTATCTACAGACTTATACCAAGCAAGTACTGG	1440	
Db	1381	TGTTGATGAGGAGTGGGAGAACTGTATCTACAGACTTATACCAAGCAAGTACTGG	1440	
Qy	1441	AAAGTGGTGTGAGGACAAACATCTATGAGAGTCTTATCAACTCTCGCGAAAGGCTTTGT	1500	
Db	1441	AAAGTGGTGTGAGGACAAACATCTATGAGAGTCTTATCAACTCTCGCGAAAGGCTTTGT	1500	
Qy	1501	AATTTGGTGAATAAATTTGAATTTCTTGGCATTTTCATCGAGCCAGTTGCGGGATAATTC	1560	
Db	1501	AATTTGGTGAATAAATTTGAATTTCTTGGCATTTTCATCGAGCCAGTTGCGGGATAATTC	1560	
Qy	1561	AGTGTGGATTTTGCATCAAGACCTGGCTTACTGCAATGAGATTAAGAGCTTTGGATGG	1620	
Db	1561	AGTGTGGATTTTGCATCAAGACCTGGCTTACTGCAATGAGATTAAGAGCTTTGGATGG	1620	
Qy	1621	TGATTTTTCGAGATCAAGATGTCGAAAATATGCTGCCACAGTGTGCTCACTTTTGG	1680	
Db	1621	TGATTTTTCGAGATCAAGATGTCGAAAATATGCTGCCACAGTGTGCTCACTTTTGG	1680	
Qy	1681	TTCTCCACAGAGACTTTGAGTGTCTTAGGCATGAGATTAAGTGAATTTCCGATGATAA	1740	
Db	1681	TTCTCCACAGAGACTTTGAGTGTCTTAGGCATGAGATTAAGTGAATTTCCGATGATAA	1740	
Qy	1741	GGTTCATGGAACCCAGTATGCTTTTCTGATGGAAATGGTAAATATCTGGTACTTTGC	1800	
Db	1741	GGTTCATGGAACCCAGTATGCTTTTCTGATGGAAATGGTAAATATCTGGTACTTTGC	1800	
Qy	1801	TCATAGAGTTGCTCAAAATGTCGCTTCAATATACCCCATCTGCTTTCCAGATTCGTTA	1860	
Db	1801	TCATAGAGTTGCTCAAAATGTCGCTTCAATATACCCCATCTGCTTTCCAGATTCGTTA	1860	
Qy	1861	TGTTGGATATAAAGTGTGTTGGTGTGATCCGGATTCATCAATGAAGTTGCTTTGAG	1920	
Db	1861	TGTTGGATATAAAGTGTGTTGGTGTGATCCGGATTCATCAATGAAGTTGCTTTGAG	1920	
Qy	1921	AAAGCATGTGCAATATGAATGATCAGACAACTAAGTAGATGCTTTGGATGGAGCA	1980	
Db	1921	AAAGCATGTGCAATATGAATGATCAGACAACTAAGTAGATGCTTTGGATGGAGCA	1980	
Qy	1981	ATATCAGCTTCTTATCTTAAATCGTCACTGATTAAGTCTTCTGTACACTTTGGAGTGA	2040	
Db	1981	ATATCAGCTTCTTATCTTAAATCGTCACTGATTAAGTCTTCTGTACACTTTGGAGTGA	2040	
Qy	2041	AGATGAAGTTCTCGAACAGCAAAAGGAGCTGTAGATCAGCTGTGCTATCTTGA	2100	
Db	2041	AGATGAAGTTCTCGAACAGCAAAAGGAGCTGTAGATCAGCTGTGCTATCTTGA	2100	
Qy	2101	TGATTTCTTTGAAGGACAGAGGCTTTTGAATTTGATGCTCTCTGGAGAGAACATAAT	2160	
Db	2101	TGATTTCTTTGAAGGACAGAGGCTTTTGAATTTGATGCTCTCTGGAGAGAACATAAT	2160	

Db	2101	TGATTCTTTGAAGGCACAGGAGGCGCTTTGGAAATTGATGCTCTCGGAGAGAACACTAATAT	2161
QY	2161	TCTCAAGGCAATGCTAAACTGTGGTTATTAAGCCCTGATGCTGAGCCCTTCTTTCAATGAT	2220
Db	2161	TCTCAAGGCAATGCTAAACTGTGGTTATTAAGCCCTGATGCTGAGCCCTTCTTTCAATGAT	2220
QY	2221	GTTGCAAACTTCCGCGCATCAAGTTGCTCGATTTCGGGACCTAGATCAAGATAATTTAT	2280
Db	2221	GTTGCAAACTTCCGCGCATCCAAGTTGCTCGATTTCGGGACCTAGATCAAGATAATTTAT	2280
QY	2281	TCCAAATGGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAATATGGTCA	2340
Db	2281	TCCAAATGGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAATATGGTCA	2340
QY	2341	GGTGTGTGTCAGTTTACTGTGTGGACATGGAGATGTTTCTGACGANTTACATCCATT	2400
Db	2341	GGTGTGTGTCAGTTTACTGTGTGGACATGGAGATGTTTCTGACGANTTACATCCATT	2400
QY	2401	TAATAACAGCAGATCCACCACAGTAATTTCATTCTCAAGGGAATGTGTTCTGCAAA	2460
Db	2401	TAATAACAGCAGATCCACCACAGTAATTTCATTCTCAAGGGAATGTGTTCTGCAAA	2460
QY	2461	AAATCCATGCTTGCATCCTGCTGTGATTAATTCGTGTTTTAAAGCGTGTAAATGTCGAGCGCT	2520
Db	2461	AAATCCATGCTTGCATCCTGCTGTGATTAATTCGTGTTTTAAAGCGTGTAAATGTCGAGCGCT	2520
QY	2521	GCACCACATGGTAGATTCTGTGTATTCCTCAGAAAGGAAGACCTCATCCGAATGA	2580
Db	2521	GCACCACATGGTAGATTCTGTGTATTCCTCAGAAAGGAAGACCTCATCCGAATGA	2580
QY	2581	ATGTTCTGGAGTGAATTTGGATGGGATATCTATTTGTTGCTGGATCAAGACATGAT	2640
Db	2581	ATGTTCTGGAGTGAATTTGGATGGGATATCTATTTGTTGCTGGATCAAGACATGAT	2640
QY	2641	CCGCGCAAGGCAGTCCAGCCGATGGATATTCCTCCAGCACCCAGCATACAGTTGACCA	2700
Db	2641	CCGCGCAAGGCAGTCCAGCCGATGGAAATTCCTCCAGCACCCAGCATACAGTTGACCA	2700
QY	2701	TGATGTCACAAATTGAGGAAGTTGAAGAGTACTTCCACCAACTATATTCTGAATCACAGTTT	2760
Db	2701	TGATGTCACAAATTGAGGAAGTTGAAGAGTACTTCCACCAACTATATTCTGAATCACAGTTT	2760
QY	2761	GGGAATCAGCAAAATGCCATGCTGTAATTCGACAGACAGAACCTCATATGCCCATGAG	2820
Db	2761	GGGAATCAGCAAAATGCCATGCTGTAATTCGACAGACAGAACCTCATATGCCCATGAG	2820
QY	2821	TGATCCATGCAAAAACCTTGCTGAGCTCTTTTCAATTGCAGTGAGCTTTCCAAAGACTGG	2880
Db	2821	TGATCCATGCAAAAACCTTGCTGAGCTCTTTTCAATTGCAGTGAGCTTTCCAAAGACTGG	2880
QY	2881	TGTTCCCGCTGAAATACCATTCTAGTTGCGCCCTAAAGAAATACCAGACTTCATGGATAA	2940
Db	2881	TGTTCCCGCTGAAATACCATTCTAGTTGCGCCCTAAAGAAATACCAGACTTCATGGATAA	2940
QY	2941	GCCGGACAGACCAGCTATATCTCAGAAGAGTTATTGGAAGGCTTTTCAGGAAGGTGAA	3000
Db	2941	GCCGGACAGACCAGCTATATCTCAGAAGAGTTATTGGAAGGCTTTTCAGGAAGGTGAA	3000
QY	3001	GGACAAAGCACCTCAGGCTAGCTCTATCCGACCTTCACAGAGATGTTTCAAGGAGATC	3060
Db	3001	GGACAAAGCACCTCAGGCTAGCTCTATCCGACCTTCACAGAGATGTTTCAAGGAGATC	3060
QY	3061	ATATGATGCTGATATGGAAGTTGATGGATTTCAGATTTACATTCCAGAAAGCTTTTGACTA	3120
Db	3061	ATATGATGCTGATATGGAAGTTGATGGATTTCAGATTTACATTCCAGAAAGCTTTTGACTA	3120
QY	3121	CAAAACTGAATATGACAAACAGCTGGTGAATTTAATGGACTACTATGGCATAAAAACAGA	3180
Db	3121	CAAAACTGAATATGACAAACAGCTGGTGAATTTAATGGACTACTATGGCATAAAAACAGA	3180
QY	3181	GGCTGAAATACTTAGTGGTGGCATTTAGAGGCATCAAAAACTTTTGACCGCAGAAAAAGA	3240
Db	3181	GGCTGAAATACTTAGTGGTGGCATTTAGAGGCATCAAAAACTTTTGACCGCAGAAAAAGA	3240

RESULT 2

US-08-100-874-1/C
 , Sequence 1, Application US/08100874
 , Patent No. 5495533
 , GENERAL INFORMATION:
 , APPLICANT: POOVAIAH, B. W.
 , APPLICANT: Takezawa, D.
 , APPLICANT: Han, T. J.
 , APPLICANT: An, G. H.
 , TITLE OF INVENTION: Control of Growth
 , TITLE OF INVENTION: Potato Plants
 , NUMBER OF SEQUENCES: 2
 , CORRESPONDENCE ADDRESS:
 , ADDRESS: Harness, Dickey & Pierce
 , STREET: P.O. Box 828
 , CITY: Bloomfield Hills
 , STATE: MI
 , COUNTRY: USA
 , ZIP: 48303
 , COMPUTER READABLE FORM:
 , MEDIUM TYPE: Floppy disk
 , COMPUTER: IBM PC compatible
 , OPERATING SYSTEM: PC-DOS/MS-DOS
 , SOFTWARE: Patent In Release #1.0, VES
 , CURRENT APPLICATION DATA:
 , APPLICATION NUMBER: US/08100.874
 , FILING DATE: July 30, 1993
 , CLASSIFICATION: 800
 , ATTORNEY/AGENT INFORMATION:
 , NAME: Smith, Dean F.
 , REFERENCE/DOCKET NUMBER: 7555-00004
 , TELECOMMUNICATION INFORMATION:
 , TELEPHONE: (313) 641-1600
 , TELEFAX: (313) 641-0270
 , TELEX: 287637 Harness UR


```
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
```

FEATURE: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:

Query Match 1.1%; Score 40; DB 4; Length 162450;
Best Local Similarity 47.6%; Pred. No. 2.1; Mismatches 0; Gaps 0;
Matches 118; Conservative 0;
QY 3032 ACCTTCACAGAGATGTTCAAGGAGATCATATGATCTGATATGGAAGTTGATGATTT 3091
Db 114325 AACCTTCAAGAGATGAACGATAAAGAGAGATGAAGATGATGAAGAAGCAGAAGAG 114384
QY 3092 GAGATTACATTCACGAAGCTTTTGACTACAAACTGATATGACACAAAGCTGGTAT 3151
Db 114385 GAGGAGGAGGAGGAAGGAAGAGGAGATGAAGATGATGACAACTGAGGAGAG 114444
QY 3152 TTAATGACTACTATGGCATAAACAGAGGCTGAAATACTTAGTGGTGCATTTGAAG 3211
Db 114445 GAGTTGAGTGCTATCCACAGGATGAAAGTCCAGTGGTATGACAGGAGGAAAT 114504
QY 3212 GATCAAAACTTTGACCCGACAAAGATGCTGAGGCCATAGCTTCTGTGAGGGCC 3271
Db 114505 CAATAATGATGAAGCTAGTATTAAGATTCGTGTCGAAGTGAGGTCCTTTAC 114564
QY 3272 TTGAGGAA 3279
Db 114565 TTGGTGCA 114572

RESULT 5
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A

REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-676-967-2
Query Match 1.0%; Score 36.8; DB 1; Length 2277;
Best Local Similarity 27.4%; Pred. No. 1.8;
Matches 113; Conservative 66; Mismatches 232; Indels 2; Gaps 1;
QY 2883 TTCGGCTGAATACCATCTCAGTTCAGTCCCTTAAGATATCCAGACTTCATGTAAGC 2942
Db 154 TTYSNATGTYNGARGAYGTNCARMGNCNTNAARGARATHACNACNTTYGARGGNTGY 213
QY 2943 CGGACACAGCAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAG 3002
Db 214 AARATHAAYGTNACNGTNGCNARARARARYTNMGNARAYARACNARAGARAGGNAAR 273
QY 3003 ACAAAGCACCACGCTAGCTCTATCGCGGACCTTCACAAAGAGATGTTGCAAGGAGATCAT 3062
Db 274 AAYGARAAYWSNGARTGYCCNARAARGARCCNAAAGCNAARAARGCNAARAGTNGCNGAY 333
QY 3063 ATGATGCTGATATGGAAGTTGATGATTTGAAGATTACATTGACGAGAGCTTTTGACTACA 3122
Db 334 AARAARGCNGNYTNATHMGNARAYTNNSNTTYARTGYWSNGAR--GAYGAYTNA 391
QY 3123 AAATGAATATGACAAAGCTGGTGAATTAATGAGGACTACTATGGCATATAAACAAGG 3182
Db 392 ARACNGTNTTGCNCARTTYGNGCNGTNYNGARGTNAAYATHCCNMGNARCCNGAY 451
QY 3183 CTGAGGCTAGTGGTGCATTTATGAAGCATCAAAACTTTTGACCGCAGAAAAGATG 3242
Db 452 GNAARATGNGGNTTYGNTTYGNTTCARTTYAARAAYTNYTNGARCGNNGNARCGNY 511
QY 3243 CTGAGGCTAGTGTGCTGCTGAGGCGCTTGAGGAAGGAGGCAAGACCGCTGG 3295
Db 512 TNAARGGNATGAAYATGAARGARATHAARGGNGMGNACNGTNGCNGTNGAYTGG 564

RESULT 6
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A

REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-098-487-2

Query Match 1.0%; Score 36.8; DB 2; Length 2277;
Best Local Similarity 27.4%; Pred. No. 1.8;
Matches 113; Conservative 66; Mismatches 232; Indels 2; Gaps 1;

QY 2883 TTCCCGCTGAATACCATCTCAGTTCGGCCCTAAAGAAATACCCAGACTTCATGGATAAGC 2942
DB 154 TTWSNATGTYNGARGAYTNCARMGNCNTYNAARGARATHACNACNTTYGARGGNTGY 213
QY 2943 CGGACAAGACCAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAG 3002
DB 214 AARATHAAYGTNACNGTNGCNAARARARYTNMGNAAYAAACNAARGAARGGNAAR 273
QY 3003 ACAAGACCTCAGGCTAGCTCTATCGGACCTTCAAGAGATGTTGCAAGGAGATCAT 3062
DB 274 AAYGARAAYWSNGARTGYCCNARAARGCCNARCCNARARCCNARARGTNGCNGAY 333
QY 3063 ATGATGCTGATATGGAAGTTGATGATTGAAGATTACATTCACCAAGCTTTTGACTACA 3122
DB 334 AARAAGCNGMNYTNATHMGNAAYTNWNTTYAARTGYWSNGAR--GAYGAYTTNA 391
QY 3123 AAACCTGAATATGACAAACAGCTGGTAAATTTAATGAGTACTATGGCATATAAAGAGAGG 3182
DB 392 ARACNGTNTTYGNCARTTYGNGCNGTNTYNGARTNAAYATHCCNMGNARCCNGAYG 451
QY 3183 CTGAAATAGTATGAGTGGCATATGAGGCATCAAAACTTTTGACCGCAGAAAAGATG 3242
DB 452 GNAARATGCMGNGNTTYGNTTYGNTTCARTTYAARAAYTYNTYNGARGCNGGNAARGCNY 511
QY 3243 CTGAGGCCATTAGTGTGTGCTGAGGCGCTTGAGGAGGAGGAGGAGGAGGAGGAGG 3295
DB 512 TNAARGGNATGAAYATGAARGARATHAARGNMGNCNGTNGCTNGAYTGG 564

RESULT 8
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435

REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-676-974-2

Query Match 1.0%; Score 36.8; DB 1; Length 2277;
Best Local Similarity 27.4%; Pred. No. 1.8;
Matches 113; Conservative 66; Mismatches 232; Indels 2; Gaps 1;

QY 2883 TTCCCGCTGAATACCATCTCAGTTCGGCCCTAAAGAAATACCCAGACTTCATGGATAAGC 2942
DB 154 TTWSNATGTYNGARGAYTNCARMGNCNTYNAARGARATHACNACNTTYGARGGNTGY 213
QY 2943 CGGACAAGACCAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAG 3002
DB 214 AARATHAAYGTNACNGTNGCNAARARARYTNMGNAAYAAACNAARGAARGGNAAR 273
QY 3003 ACAAGACCTCAGGCTAGCTCTATCGGACCTTCAAGAGATGTTGCAAGGAGATCAT 3062
DB 274 AAYGARAAYWSNGARTGYCCNARAARGCCNARCCNARARCCNARARGTNGCNGAY 333
QY 3063 ATGATGCTGATATGGAAGTTGATGATTGAAGATTACATTCACCAAGCTTTTGACTACA 3122
DB 334 AARAAGCNGMNYTNATHMGNAAYTNWNTTYAARTGYWSNGAR--GAYGAYTTNA 391
QY 3123 AAACCTGAATATGACAAACAGCTGGTAAATTTAATGAGTACTATGGCATATAAAGAGAGG 3182
DB 392 ARACNGTNTTYGNCARTTYGNGCNGTNTYNGARTNAAYATHCCNMGNARCCNGAYG 451
QY 3183 CTGAAATAGTATGAGTGGCATATGAGGCATCAAAACTTTTGACCGCAGAAAAGATG 3242
DB 452 GNAARATGCMGNGNTTYGNTTYGNTTCARTTYAARAAYTYNTYNGARGCNGGNAARGCNY 511
QY 3243 CTGAGGCCATTAGTGTGTGCTGAGGCGCTTGAGGAGGAGGAGGAGGAGGAGGAGG 3295
DB 512 TNAARGGNATGAAYATGAARGARATHAARGNMGNCNGTNGCTNGAYTGG 564

RESULT 7
US-09-098-487-2
Sequence 2, Application US/09098487
Patent No. 5917025
GENERAL INFORMATION:
APPLICANT: COLLINS, Kathleen
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,487
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A

APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: No. 5348582el Prokaryotic Polynucleotides,
TITLE OF INVENTION: Polypeptides and their Uses
NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936.165A
FILING DATE: 24-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/027,032
FILING DATE: 24-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimml, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 3191 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-936-165A-194

Query Match 1.0%; Score 36.4; DB 4; Length 3191;
Best Local Similarity 49.0%; Pred. No. 2.8;
Matches 94; Conservative 0; Indels 0; Gaps 0;

QY 260 TTACAGAAATATACAGATATGGAAGTTCATTTGCCGACAAATGCTGCTGACAAATATC 319
DB 2823 TTAAGAAACACATATTTATGAAGATGTTTGACACCGGAGATATTGAAAAATTTAT 2882
QY 320 GGATCTAGACATTTGCCAAAGTTCAATTTGCCGACAAATGCTGCTGACAAATATC 379
DB 2883 CGNCTAATCGTGGTCAATATATGTTGTTGCGAGATAAAAGAAACAAAGGATTT 2942
QY 380 ACTTTGGCTAATACAGGCTGATTTTGGCTCTCTTATTTCAAGGCTTGGGAAATGAA 439
DB 2943 ACCTTTCCCTAAAGAAAGTCAGTATTTTGAACACTTGTACTTGTAGTGGATCAGNRAAT 3002
QY 440 ACTGATATTGTC 451
DB 3003 CCTGGTGGTGGC 3014

RESULT 10
US-09-415-946-1/c
Sequence 1, Application US/09415946
Patent No. 6376751
GENERAL INFORMATION:
APPLICANT: Sung, Z. Renee
APPLICANT: Aubert, Dominique
APPLICANT: Chen, Lingjing
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Nucleic Acids That Control Reproductive Development in
TITLE OF INVENTION: Plants

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935.313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 1.0%; Score 36.6; DB 1; Length 7218;
Best Local Similarity 10.1%; Pred. No. 3.8; Indels 0; Gaps 0;
Matches 48; Conservative 204; Mismatches 233; Indels 0; Gaps 0;

QY 581 TTTTCTTTCTTTATCTAGTGTCTGCTGACTATTAACCTCAGCTTTCATATGAAAT 640
DB 1074 YY 1133
QY 641 ATATGCGAGTGTGCTCATGCTCATATGCTCAAAATGCTCAGTTCTCTCATACAG 700
DB 1134 YY 1193
QY 701 TTATTTGGTCTCGGATCTATAGAGACTGAAACTCTCTGTATAGCTTCTTTAAG 760
DB 1194 YY 1253
QY 761 GAACCTCGATGATCAGTGGGTGAGGACAAACAGATTTCCCTCCATCTTGGATAGGCTA 820
DB 1254 YY 1313
QY 821 TCTCTAGCTATGTTTGGAGTTCCTGAGGCTGCTGCTTCCAAATTCAGGAAGT 880
DB 1314 YY 1373
QY 881 TTCTTCCATCATCAGACGTGAAACATATATCTTTACAGACTGGTTTCACTTTTC 940
DB 1374 YY 1433
QY 941 GTCTCTAAAATCGGCTCTGTTCCCATGCTCAGGCTCCGGAAGGAATTCATTCCTC 1000
DB 1434 YGTACCAATTTCTCTATCTCTTTAACTACTTGTGATAGTAGTAAATACAGTAGGCT 1493
QY 1001 TACAAGATTTTGTTCAAAATAGTCTTCTTGTGACAGCAATGATGATCCTGGC 1055
DB 1494 ACATGCGGTTTTTGAACGATAGATGCTGCTAGAACGATGCTACGCTAGTC 1548

RESULT 9
US-08-936-165A-194
Sequence 194, Application US/08936165A
Patent No. 6348582
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lonetto, Michael
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Richard

FILE REFERENCE: 018941-000110US
CURRENT APPLICATION NUMBER: US/09/415,946
CURRENT FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 09/169,696
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 17341
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: genomic DNA (Ecotype Columbia) from CD82 clone
NAME/KEY: Promoter
LOCATION: (1)..(3201)
NAME/KEY: CDS
LOCATION: Join(4241..4335, 4448..4623, 4704..4823, 4903..4956,
OTHER INFORMATION: EMBRYONIC FLOWER 1 (EMF1)
NAME/KEY: exon
LOCATION: (3202)..(3265)
OTHER INFORMATION: exon 1
NAME/KEY: intron
LOCATION: (3266)..(4159)
NAME/KEY: exon
LOCATION: (4160)..(4335)
OTHER INFORMATION: exon 2
NAME/KEY: intron
LOCATION: (4336)..(4447)
NAME/KEY: exon
LOCATION: (4448)..(4523)
OTHER INFORMATION: exon 3
NAME/KEY: intron
LOCATION: (4524)..(4703)
NAME/KEY: exon
LOCATION: (4704)..(4823)
OTHER INFORMATION: exon 4
NAME/KEY: intron
LOCATION: (4824)..(4902)
NAME/KEY: exon
LOCATION: (4903)..(4956)
OTHER INFORMATION: exon 5
NAME/KEY: intron
LOCATION: (4957)..(5045)
NAME/KEY: exon
LOCATION: (5046)..(6307)
OTHER INFORMATION: exon 6
NAME/KEY: intron
LOCATION: (6308)..(6447)
NAME/KEY: exon
LOCATION: (6448)..(8065)
OTHER INFORMATION: exon 7
NAME/KEY: intron
LOCATION: (8066)..(8300)
NAME/KEY: exon
LOCATION: (8301)..(8648)
OTHER INFORMATION: exon 8
US-09-415-946-1

Query Match 0.9%; Score 35.4; DB 4; Length 17341;
Best Local Similarity 49.2%; Pred. No. 13;
Matches 93; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1115 ATGAGCATGCTTAGAAGAACTGTACTATATAAAGGAGTGCTGTATGATGCCGTGAGG 1174
DB 10360 ATTTGACCTAACTTTGTAATTTCTTGAAGTAAAGAACAGTGTAGATCTTAAGAAGATG 10301

QY 1175 TGGCTCACTAGCAGTAGTGGTATCTCAAGGTAGAACCTCCAAATCTCCGTCC 1234
DB 10300 TTCTCTAGTGGTCTGATGAGATCTCTTAGAACAGAGCTTCCAAGTATGACGGA 10241

QY 1235 ATCACTTTAGATGGGTGGTGTATGTAAAGAGGTCTCTAGTAACACCATGCAAGTT 1294
DB 10240 TGTGCTTGAAGAAATGTGTGTGTCTTCAAGAAAGACATTTCAACCTTCCAAAGT 10181

QY 1295 TATTTTGT 1303
DB 10180 TACTTTGT 10172

RESULT 11
US-09-134-001C-342
Sequence 342, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 342
LENGTH: 1248
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-342

Query Match 0.9%; Score 35; DB 4; Length 1248;
Best Local Similarity 51.6%; Pred. No. 4.3;
Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 3040 AAGAGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGATTGAAGATTA 3099
DB 813 ATGCTTTTACAGGTAGCTTAGAAATGACCATGGAGATGCAATGAAGTGTAGATT 872
QY 3100 CATTGACGAAGCTTTTGACTACAAACTGAATATGACACAAAGCTGGTGAATTATGGA 3159
DB 873 GCTAAACAGCCTTAGAAGTCGATATGATATCAACCACTGTGTTTAACTGA 932

QY 3160 CTACTATGGCATAAACACAGAGCTGAAATACTTA 3194
DB 933 TTATATCGTAATGAAGAGACTATGATCAATTA 967

RESULT 12
US-08-911-445-4
Sequence 4, Application US/08911445
Patent No. 5876713
GENERAL INFORMATION:
APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.
TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE
TITLE OF INVENTION: AMIDOTRANSFERASE, ITS PRODUCTION AND USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: FETHERSTONHAUGH & Co.
STREET: P.O. BOX 2999, STATION D
CITY: OTTAWA
STATE: ONTARIO
COUNTRY: CANADA
ZIP: K1P 5Y6
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,445
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 213944-1996
FILING DATE: 13-AUG-1996

Query Match	Best Local Similarity	Score	DB 2	Length	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB
-------------	-----------------------	-------	------	--------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	----

```

Query Match      0.9%; Score 35; DB 4; Length 1458;
Best Local Similarity 52.4%; Pred. No. 4.7;
Matches 77; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 3130 ATATCACAAACGCTGGGTAAATTTAATGGACTACTATGCATATAAAACAGAGGCTGAAAT 3189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 372 AAATTACAAAGNCTTGGGAATTTCTGGAAGCAAGGCTACGAGTTTGAGTCAGAAAC 431
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3190 ACTTTAGTGGTCATTATGAAGGCATCAAAACTTTTGACCCGCAAAAGATGCTTGAGGC 3249
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 432 AGATACAGAGACCATCGGCAAGCTGATTAAATATGTGTTGCAACACAGAGAAACTGAGGA 491
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3250 CATTAGTGTGCTGTGAGGGCCTTGAG 3276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 492 CATTACGTTTCAACGTTGGTCGAGAG 518
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: November 5, 2002, 23:30:56
Job time : 710 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 22:34:33 ; Search time 9169 Seconds
(without alignments)
11842.357 Million cell updates/sec

Title: US-09-782-874-1
Perfect score: 3731
Sequence: 1 GAAATATCTTACTTACTT.....AGTTTCATCTTCTTCTTAAA 3731

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_hcg_hum.*
- 31: em_hcg_inv.*
- 32: em_hcg_other.*
- 33: em_hcg_mus.*
- 34: em_hcg_pln.*
- 35: em_hcg_rod.*
- 36: em_hcg_mam.*
- 37: em_hcg_vrt.*
- 38: em_sy.*
- 39: em_hgtgo_hum.*
- 40: em_hgtgo_mus.*
- 41: em_hgtgo_other.*

SD

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3731	100.0	3731	6	ARI45904	ARI45904 Sequence
2	3680	98.6	3731	8	LERDRP	Y10403 L.esculentu
3	94	2.5	3505	8	NTAC11576	AJ011576 Nicotiana
4	73	2.0	4579	8	PHY011979	AJ011979 petunia h
5	35	0.9	1838	8	ATH011977	AJ011977 Arabidops
6	30	0.8	908	6	IL8797	IL8797 Sequence 1
7	30	0.8	906	6	POTCAM	J04559 Solanum tub
8	25	0.7	25	6	ARI45905	ARI45905 Sequence
9	25	0.7	25	6	ARI45907	ARI45907 Sequence
10	24	0.6	112247	25	ARI45906	ARI45906 Sequence
11	23	0.6	132699	8	AC095524	AC095524 Rattus no
12	23	0.6	132699	8	AC006917	AC006917 Genomic s
13	22	0.6	204160	2	AC113586	AC113586 Mus muscu
14	21	0.6	54683	2	AL732358_3	Continuation (4 of
15	21	0.6	62512	2	AC103999	AC103999 Homo sapi
16	21	0.6	64179	9	AL627224	AL627224 Human DNA
17	21	0.6	86478	2	AC022577	AC022577 Homo sapi
18	21	0.6	101706	2	AP004079	AP004079 Oryza sat
19	21	0.6	122685	2	AC099102	AC099102 Rattus no
20	21	0.6	129226	9	HS960017	AL022166 Human DNA
21	21	0.6	131310	2	AC126008	AC126008 Medicago
22	21	0.6	135458	2	AC117166	AC117166 Rattus no
23	21	0.6	137723	9	HS060622	AL034551 Human DNA
24	21	0.6	138931	2	AP004023	AP004023 Oryza sat
25	21	0.6	142234	9	AC110620	AC110620 Homo sapi
26	21	0.6	150055	2	AC108260	AC108260 Rattus no
27	21	0.6	153258	2	AC128389	AC128389 Rattus no
28	21	0.6	153859	2	AC121213	AC121213 Rattus no
29	21	0.6	155904	2	AC015953	AC015953 Homo sapi
30	21	0.6	157759	2	AC114071	AC114071 Rattus no
31	21	0.6	158158	2	AC034304	AC034304 Homo sapi
32	21	0.6	158371	9	AC011155	AC011155 Homo sapi
33	21	0.6	160083	2	AC129367	AC129367 Rattus no
34	21	0.6	162600	2	AC091357	AC091357 Rattus no
35	21	0.6	165699	2	AP001802	AP001802 Homo sapi
36	21	0.6	168763	2	AC095109	AC095109 Rattus no
37	21	0.6	171196	2	AP001369	AP001369 Homo sapi
38	21	0.6	171853	9	AP000880	AP000880 Homo sapi
39	21	0.6	175245	2	AC126420	AC126420 Mus muscu
40	21	0.6	177034	9	AP002840	AP002840 Homo sapi
41	21	0.6	178228	2	AC009420	AC009420 Homo sapi
42	21	0.6	179025	2	AC109179	AC109179 Mus muscu
43	21	0.6	179281	2	AP001934	AP001934 Homo sapi
44	21	0.6	181265	2	AC083994	AC083994 Homo sapi
45	21	0.6	183951	2	AC113269	AC113269 Mus muscu
46	21	0.6	185148	9	AC096641	AC096641 Homo sapi
47	21	0.6	186386	2	AC084367	AC084367 Homo sapi
48	21	0.6	189718	2	AC125777	AC125777 Rattus no
49	21	0.6	193279	9	AC094466	AC094466 Homo sapi
50	21	0.6	202781	2	AC055817	AC055817 Mus muscu

ALIGNMENTS

RESULT 1
ARI45904
LOCUS
DEFINITION Sequence 1 from patent US 6218142.
ACCESSION ARI45904
VERSION ARI45904.1 GI:15109093
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

ARI45904 3731 bp DNA 1linear PAT 08-AUG-2001

Pred. No. is the number of results predicted by chance to have a

REFERENCE 1 (bases 1 to 3731)
AUTHORS Wassenegeger,M., Riedel,L., Schiebel,W. and Sanger,H.L.
TITLE Nucleic acid molecules encoding polypeptides having the enzymatic activity of an RNA-directed RNA polymerase (RDRP)
JOURNAL Patent: US 6218142-A 1 17-APR-2001;
FEATURES Location/Qualifiers
source 1..3731
BASE COUNT 1064 a 569 c 849 g 1149 t
ORIGIN
Query Match 100.0%; Score 3731; DB 6; Length 3731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATAATTCCTTACTTACACAGGATGACTCATCACTCCCTCAAGTCCTTTGTGT 60
Db 1 GAATAATTCCTTACTTACACAGGATGACTCATCACTCCCTCAAGTCCTTTGTGT 60
QY 61 GTTGATATAAATAATTTGGTTGTCTTCAGTTTCAGTCACACTACTGCTGGGTAGTTTTAT 120
Db 61 GTTGATATAAATAATTTGGTTGTCTTCAGTTTCAGTCACACTACTGCTGGGTAGTTTTAT 120
QY 121 TTTCGATAAATTCAGGGGTATTCAGATGGTGTAGCATTTGAAAGTCGAATGCACCT 180
Db 121 TTTCGATAAATTCAGGGGTATTCAGATGGTGTAGCATTTGAAAGTCGAATGCACCT 180
QY 181 GGAATTTGGCTACATGGGAAGACAATTCAGGTTTTCGGATTCCTTATCTCTCTGC 240
Db 181 GGAATTTGGCTACATGGGAAGACAATTCAGGTTTTCGGATTCCTTATCTCTCTGC 240
QY 241 GGAAGTGGTTAAGTCATCTTTAGAGAATATACAGGATATGGAACCTCTATGTCATTTGA 300
Db 241 GGAAGTGGTTAAGTCATCTTTAGAGAATATACAGGATATGGAACCTCTATGTCATTTGA 300
QY 301 GGTTAAACAGTCCAAAGGAGATCAGACATTTGCCAAAGTCAATTTGCCGACACAT 360
Db 301 GGTTAAACAGTCCAAAGGAGATCAGACATTTGCCAAAGTCAATTTGCCGACACAT 360
QY 361 AAGTGTGACAAAATCATCACTTTGGCTAATAACAGGCTGTATTTGGCTCTCTTATTT 420
Db 361 AAGTGTGACAAAATCATCACTTTGGCTAATAACAGGCTGTATTTGGCTCTCTTATTT 420
QY 421 GAAGCTTTGGGAATGAACCTGATATTTGCCAACTCGGGCATATGTCGATCAGATGGA 480
Db 421 GAAGCTTTGGGAATGAACCTGATATTTGCCAACTCGGGCATATGTCGATCAGATGGA 480
QY 481 TGGCATACCTTTGAATTCGGATGTCAGATATCAGATGACAAGTTTGCAGTTTGGGAAG 540
Db 481 TGGCATACCTTTGAATTCGGATGTCAGATATCAGATGACAAGTTTGCAGTTTGGGAAG 540
QY 541 TACAGAAGTTTCAATTCATTTGGCATTTGGATTGAAGAAATTTTTTCTTTTATCTAG 600
Db 541 TACAGAAGTTTCAATTCATTTGGCATTTGGATTGAAGAAATTTTTTCTTTTATCTAG 600
QY 601 TGGTTACGCTACATATAACTTCAGCTTTCATATGAAATATATGCGAGCTTGTGCTCCA 660
Db 601 TGGTTACGCTACATATAACTTCAGCTTTCATATGAAATATATGCGAGCTTGTGCTCCA 660
QY 661 TCCTCCATATGGTCAAAATGCTAGTTTCTCCCTCATACAGTTATTTGGTCTCTCTGGAT 720
Db 661 TCCTCCATATGGTCAAAATGCTAGTTTCTCCCTCATACAGTTATTTGGTCTCTCTGGAT 720
QY 721 CTATAAGAGACTTGAACCTCCGTTATAGCTTCTTTAAGAAACCTCTGATCAGTG 780
Db 721 CTATAAGAGACTTGAACCTCCGTTATAGCTTCTTTAAGAAACCTCTGATCAGTG 780
QY 781 GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGCTATCTCTAGCTTATGTTTGA 840
Db 781 GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGCTATCTCTAGCTTATGTTTGA 840
QY 841 GTTCCGTAGGGTGTCTGCTTCCAAATTTTCAGGAAGTTTTTTCACATATCAGAACG 900
Db 841 GTTCCGTAGGGTGTCTGCTTCCAAATTTTCAGGAAGTTTTTTCACATATCAGAACG 900

Db 841 GTTCCGTAGGGTGTCTGCTTCCAAATTTTCAGGAAGTTTTTTCACATATCAGAACG 900
QY 901 TGAACAAATATTTACTTTACAGACTGGTTTCACTTTTTCGCTCTCAAAAATCGGCTCT 960
Db 901 TGAACAAATATTTACTTTACAGACTGGTTTCACTTTTTCGCTCTCAAAAATCGGCTCT 960
QY 961 GGTTCCTCAATGCTCCAGCTCCGGAAGAAATTTCAATTCCTCAAGATTTTGTTCAAAAT 1020
Db 961 GGTTCCTCAATGCTCCAGCTCCGGAAGAAATTTCAATTCCTCAAGATTTTGTTCAAAAT 1020
QY 1021 TACTTCTTTTGGTACAGCATGGATGATACCTGGCCAGCATTAATAGTCTACTTTTTCCG 1080
Db 1021 TACTTCTTTTGGTACAGCATGGATGATACCTGGCCAGCATTAATAGTCTACTTTTTCCG 1080
QY 1081 ATTAGTGTATCTCGAAGGAGAAATGTGGCATGCTAGAGCATGCTTATAGAGAACTGTA 1140
Db 1081 ATTAGTGTATCTCGAAGGAGAAATGTGGCATGCTAGAGCATGCTTATAGAGAACTGTA 1140
QY 1141 CTATATAAGAGAGTGTCTTATGATCCGTTAGGTGGTCTGAGCAGTATGATGGTGA 1200
Db 1141 CTATATAAGAGAGTGTCTTATGATCCGTTAGGTGGTCTGAGCAGTATGATGGTGA 1200
QY 1201 TCTCAAGGGTAGACAACCTCCAAAATCTCCGTCATCACTTTAGATGGGTGGTGA 1260
Db 1201 TCTCAAGGGTAGACAACCTCCAAAATCTCCGTCATCACTTTAGATGGGTGGTGA 1260
QY 1261 TGTAAAGAGGTCCTAGTAAACCATGCAAAATTTATTTTGTGTCAGAGAGTTAAATGT 1320
Db 1261 TGTAAAGAGGTCCTAGTAAACCATGCAAAATTTATTTTGTGTCAGAGAGTTAAATGT 1320
QY 1321 TTCCAATCGGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTTCGTTCTTTT 1380
Db 1321 TTCCAATCGGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTTCGTTCTTTT 1380
QY 1381 TGTGTAGAGGAGTGGGAGAACTGTTTCTACAGACTTATTACCAAAAGCAAGTACTGG 1440
Db 1381 TGTGTAGAGGAGTGGGAGAACTGTTTCTACAGACTTATTACCAAAAGCAAGTACTGG 1440
QY 1441 AAGTGTGTGTCAGACAAACATCTATGAGAGGATCTTATCAACTCTGCGAAAAGGTTGT 1500
Db 1441 AAGTGTGTGTCAGACAAACATCTATGAGAGGATCTTATCAACTCTGCGAAAAGGTTGT 1500
QY 1501 AATTGTGATATAAAATTTTGAATTTCTTCGATTTTCATCGAGCCAGTTGCGGGATAATTC 1560
Db 1501 AATTGTGATATAAAATTTTGAATTTCTTCGATTTTCATCGAGCCAGTTGCGGGATAATTC 1560
QY 1561 AGTGTGGATTTGTCATCAAGACCTGGCCCTTACTGCAATGATATAAGAGCTTGGATGG 1620
Db 1561 AGTGTGGATTTGTCATCAAGACCTGGCCCTTACTGCAATGATATAAGAGCTTGGATGG 1620
QY 1621 TGATTTTTCGAGATCAAGATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTTGG 1680
Db 1621 TGATTTTTCGAGATCAAGATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTTGG 1680
QY 1681 TTCCTCCAGAGACTTTGAGTGTCTTAGGATGAGATTGAAGTTATTCCTCGATGTA 1740
Db 1681 TTCCTCCAGAGACTTTGAGTGTCTTAGGATGAGATTGAAGTTATTCCTCGATGTA 1740
QY 1741 GGTTCATGAACACAGCTATGCTTTTCTGATGGAATTTGGTAAATATCTGTGATCTTGC 1800
Db 1741 GGTTCATGAACACAGCTATGCTTTTCTGATGGAATTTGGTAAATATCTGTGATCTTGC 1800
QY 1801 TCATAGAGTTGGCTCAAAATGTCGCTTCAATATACCCCATCTCTTCCAGATCTGTA 1860
Db 1801 TCATAGAGTTGGCTCAAAATGTCGCTTCAATATACCCCATCTCTTCCAGATCTGTA 1860
QY 1861 TGGTGGATATAAGGTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1920
Db 1861 TGGTGGATATAAGGTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1920
QY 1921 AAAGAGCATGTGCAATATGATATCAGACAAACATAAAGTTAGATGTCTTGGATGGAGCA 1980
Db 1921 AAAGAGCATGTGCAATATGATATCAGACAAACATAAAGTTAGATGTCTTGGATGGAGCA 1980

QY 1981 ATATCAGCCTTGTTATCTTAATCGTCAACTGATTAAGCTCTTGCTACACTTGGAGTAA 2040
DB 1981 ATATCAGCCTTGTTATCTTAATCGTCAACTGATTAAGCTCTTGCTACACTTGGAGTAA 2040
QY 2041 AGATGAAGTCTCGACAGAGCAAAAGNAAGCTGTAGATCAGCTTGATCTATCTTGA 2100
DB 2041 AGATGAAGTCTCGACAGAGCAAAAGNAAGCTGTAGATCAGCTTGATCTATCTTGA 2100
QY 2101 TGAATCTTTGAAGCACAGGAGCTTTGGAATGTATGCTCTCGAGAGAACACTAATAT 2160
DB 2101 TGAATCTTTGAAGCACAGGAGCTTTGGAATGTATGCTCTCGAGAGAACACTAATAT 2160
QY 2161 TCTCAAGGCAATGCTAAACTGTGGTTATTAAGCCTGATGCTGAGCCCTTTCTTTCAATGAT 2220
DB 2161 TCTCAAGGCAATGCTAAACTGTGGTTATTAAGCCTGATGCTGAGCCCTTTCTTTCAATGAT 2220
QY 2221 GTTGCAAACTTCGCGCATCCAAAGTGTCTCGATTTGCGGACTAGATCAAGAATATTTAT 2280
DB 2221 GTTGCAAACTTCGCGCATCCAAAGTGTCTCGATTTGCGGACTAGATCAAGAATATTTAT 2280
QY 2281 TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGATATGGTCA 2340
DB 2281 TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGATATGGTCA 2340
QY 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTGTTCGAGATTTTACATCCCAT 2400
DB 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTGTTCGAGATTTTACATCCCAT 2400
QY 2401 TAATAACAGCAGATCCACCAACAGTAATTCATTCCTGAAGGGAATGTGGTTGTTGCAAA 2460
DB 2401 TAATAACAGCAGATCCACCAACAGTAATTCATTCCTGAAGGGAATGTGGTTGTTGCAAA 2460
QY 2461 AATCCATGCTTCATCTCTGCTGATTCGTTTAAAGGCTGTAATGTTCGAGCGT 2520
DB 2461 AATCCATGCTTCATCTCTGCTGATTCGTTTAAAGGCTGTAATGTTCGAGCGT 2520
QY 2521 GCACCATGTTGATGTTGTTATTCCTCAGAAAGGAAAGAACCTTCATCCGAATGA 2580
DB 2521 GCACCATGTTGATGTTGTTATTCCTCAGAAAGGAAAGAACCTTCATCCGAATGA 2580
QY 2581 ATGTTCTGGAGTATGATGGGATATCTACTTTGTTGCTGGGATCAAGACATGAT 2640
DB 2581 ATGTTCTGGAGTATGATGGGATATCTACTTTGTTGCTGGGATCAAGACATGAT 2640
QY 2641 CCGCCCAAGCAAGTCCAGCCGATGAATATCTCCAGCAGCAGACATACAGTTGGACCA 2700
DB 2641 CCGCCCAAGCAAGTCCAGCCGATGAATATCTCCAGCAGCAGACATACAGTTGGACCA 2700
QY 2701 TGAATGCAATGAGGAAGTTGAAGAGTACTTCACTCACTATATGTGAATGACAGTTT 2760
DB 2701 TGAATGCAATGAGGAAGTTGAAGAGTACTTCACTCACTATATGTGAATGACAGTTT 2760
QY 2761 GGAATCATAGCAAAATGCCCATGCTCTATTTGCAGACAGAACCTGATATGCCATGAG 2820
DB 2761 GGAATCATAGCAAAATGCCCATGCTCTATTTGCAGACAGAACCTGATATGCCATGAG 2820
QY 2821 TGATCCATGCAAAAACTTGCTGAGCTCTTTTCAATGTCAGTGGACTTTCCAAAGACTGG 2880
DB 2821 TGATCCATGCAAAAACTTGCTGAGCTCTTTTCAATGTCAGTGGACTTTCCAAAGACTGG 2880
QY 2881 TGTTCCTGCTGAATACCAATCTCAGTTGGGCTTAAGAAATACCCAGACTTCATGATGAA 2940
DB 2881 TGTTCCTGCTGAATACCAATCTCAGTTGGGCTTAAGAAATACCCAGACTTCATGATGAA 2940
QY 2941 GCGGCAAGACAGCAGTATATCTCAGAAAGAGTTATGTGAAAGCTTTTCCAGAAAGTGA 3000
DB 2941 GCGGCAAGACAGCAGTATATCTCAGAAAGAGTTATGTGAAAGCTTTTCCAGAAAGTGA 3000
QY 3001 GGCACAAAGCACCTCAGGCTAGCTCTATTCGCGACCTTTCACAGAGATGTTGCAAGGATC 3060
DB 3001 GGCACAAAGCACCTCAGGCTAGCTCTATTCGCGACCTTTCACAGAGATGTTGCAAGGATC 3060

QY 3061 ATATGATGCTGATATGGAAGTTGATGATGATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120
DB 3061 ATATGATGCTGATATGGAAGTTGATGATGATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120
QY 3121 CAAAACTGAATATGACAAAGCTGGGTAAATTAATGACTACTATGCAATAAAACAGA 3180
DB 3121 CAAAACTGAATATGACAAAGCTGGGTAAATTAATGACTACTATGCAATAAAACAGA 3180
QY 3181 GGCTGAAATACTTTAGTGTGGCATTATGAAGGCATCAAAAACTTTTGACCCGAGAAAAGA 3240
DB 3181 GGCTGAAATACTTTAGTGTGGCATTATGAAGGCATCAAAAACTTTTGACCCGAGAAAAGA 3240
QY 3241 TGCTGAGGCCATTAGTGTGCTGTGAGGCCCTTGAGGAAGGAGGCAAGAGCTGGTTCAA 3300
DB 3241 TGCTGAGGCCATTAGTGTGCTGTGAGGCCCTTGAGGAAGGAGGCAAGAGCTGGTTCAA 3300
QY 3301 GAGGCGTAATGATATGATGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTACATA 3360
DB 3301 GAGGCGTAATGATATGATGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTACATA 3360
QY 3361 TCATCCTACATATTTGGGTTGCTACAAATCAGGGGTTGAAAAGAGCTCAITTTTCATTAGCTT 3420
DB 3361 TCATCCTACATATTTGGGTTGCTACAAATCAGGGGTTGAAAAGAGCTCAITTTTCATTAGCTT 3420
QY 3421 TCCCTGGTGTGTTATGACCAAGCTAATCCAGATTAAAGAGGACAAAGCACGTAACAGGCC 3480
DB 3421 TCCCTGGTGTGTTATGACCAAGCTAATCCAGATTAAAGAGGACAAAGCACGTAACAGGCC 3480
QY 3481 AGTCTCAACTTGTCATCTCAGGGCTCAACTGAGTACAGATTAGTGTGAAATGAGA 3540
DB 3481 AGTCTCAACTTGTCATCTCAGGGCTCAACTGAGTACAGATTAGTGTGAAATGAGA 3540
QY 3541 TTCCAGTCGAGCGTTAAGCTGAT 3600
DB 3541 TTCCAGTCGAGCGTTAAGCTGAT 3600
QY 3601 TGTATGCAATTTTGACTACCTTTTGTCTTTAAACTGCATGAAGCTGCACATATATGTC 3660
DB 3601 TGTATGCAATTTTGACTACCTTTTGTCTTTAAACTGCATGAAGCTGCACATATATGTC 3660
QY 3661 AGTACTCAAGAACAGATGTACAGCTAAGTAAATATATATATATATATATATATATATAT 3720
DB 3661 AGTACTCAAGAACAGATGTACAGCTAAGTAAATATATATATATATATATATATATATAT 3720
QY 3721 TTTCTTCTAAA 3731
DB 3721 TTTCTTCTAAA 3731
RESULT 2
LERDRP
LOCUS LERDRP L.esculentum mRNA for RNA-directed RNA polymerase. 3731 bp mRNA linear PLN 18-DEC-1998
DEFINITION
ACCESSION Y10403
VERSION Y10403.1 GI:4038591
KEYWORDS RDRP gene; RNA-directed RNA polymerase.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 3731)
AUTHORS Schiebel,W., Pelissier,T., Riedel,L., Thalmair,S., Schiebel,R.,
Kempe,D., Lottspeich,F., Sanger,H.L. and Wassenecker,M.
Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from tomato
JOURNAL Plant Cell 10 (12), 2087-2101 (1998)
MEDLINE 98055198
PUBMED 9836747
REFERENCE 2 (bases 1 to 3731)
AUTHORS Wassenecker,M.
TITLE Direct Submission

JOURNAL

Submitted (08-JAN-1997) M. Wassenegger, Max-Planck-inst. fuer
Biochemie, Viroidforschung, Am Klopferspitz 18a,
Planegg-Martinsried, 82152, FRG

FEATURES

Source

Location/Qualifiers

1..3731

/organism="Lycopodium obscurum"

/strain="Rutgers"

/db_xref="taxon:4081"

194..3538

/gene="RdRp"

194..3538

/gene="RdRp"

/EC_number="2.7.7.48"

/codon_start=1

/product="RNA-directed RNA polymerase"

/protein_id="CAA71421.1"

/db_xref="GI:4038592"

/translation="MGKTIQVGFVPLLSAEVVKFLEKTYGTGVCALAEVKQSKGGS

RAFAQVQADNTSADKIITLANNRLVFGSSYLKAMEMKTDIVLEAYVDWDGTLNP

GCOISDDKPAVIGSTBVSIOFGIGIKKPEFLSSGSADYKKQLSVENWVYLHPYV

ONAOPLLQLGNAPRIYKLENSCSFKEFEDDQVNRITDFPSWIGLSSSLCLQPR

RGVLPNFEEFPHARENNITLQGTFFVSKSALVPNPPEGISIPYKILFKI

SSLVQHGCIIPGALNVPFRLVDRNRNVACIEHALEKLYIKKCCYDPVRLWTEQYD

GYLKGQPKPSITLDDGLVVRVLPVCKVYFCGPEVNVSNVLENYSEDIDNFL

RVFVEDEMKLYSTDLLPKASTGSTGVRNIVYERILSLRKGFVIGDKKFLAFSS

QLRDNVMEFASRPGTLANDIRAMGDFSOIKNAVYARLQGVGGSSRGLSVLRHE

IEVTPDKVFGTSYVSDGIGIKSGDFAHVASKGLOVTPSAFOIRYGVYGVGVVD

PDSMKLSLRKMSKVESDNIKLVLGHSKQPCVNLRLITLILSLGVKVELEQKQ

KEAVDQDAILHDSLKAQELSELSPGENTNLKALNCGKPKPAEPFLSMQLQFRA

SKLQDTRSRIFVNGRTMGCLESTRELVYGVQVFTGAGHGFSDDDLHPFNNS

STNSNLIKNNVAVKPNCLHPGDIRLVKANVRALHMDVCPVFPQKGRPHNES

GSLLDQDIFVCDQDIMPQVPMQPPAPSIQLDHDTIEEVEEFTNYIVNDSL

GIITANHVFADEPDMASDPCKLAELFSLAVDFPKTGVPAEIPSLRKPKEYPDM

DKPKTYSISERVIGLKRKVDKAPQASSIATFTRDVARRSYDADMEVDGFEDYIDE

APDYKTEYNLQNLMDYVGIKTEAELISGLGIMKASKYEDRDKDAEALSVAVRLRE

ARAFKERNIDMDLPLKASAWVHYVHTPYWGCYNQGLKRAHFISSFPCVYDQLIQIK

KDKARNPVLNLSLRAQLSHRLVLK"

BASE COUNT 1064 a 688 c 850 g 1149 t

ORIGIN

Query Match

Best Local Similarity 98.6%; Score 3680; DB 8; Length 3731;

Mismatches 0; Conservative 0; Gaps 0;

Mismatches 3730; Indels 1; Indels 0; Gaps 0;

1 GAATATCTTCTTACTTACCTCACCAGGATGACTCATCACCCTCCCTCAAGCTTTTGT 60

1 GAATATCTTCTTACTTACCTCACCAGGATGACTCATCACCCTCCCTCAAGCTTTGT 60

61 GTTGTGATAAATAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGGTAGTTTAT 120

61 GTTGTGATAAATAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGGTAGTTTAT 120

121 TTTGCATAACTTCAGGGGATTCAGTGTGTTAGCATTTGAAGTCAAGTCACTT 180

121 TTTGCATAACTTCAGGGGATTCAGTGTGTTAGCATTTGAAGTCAAGTCACTT 180

181 GGAATTTGGTGTACATGGGAAGACAATTCAGTTTTCGGATTCCTTCTCTG 240

181 GGAATTTGGTGTACATGGGAAGACAATTCAGTTTTCGGATTCCTTCTCTG 240

241 GGAAGTGGTAAAGTCATCTTAGAGAAATATACAGATATGGAATGTATGTGATGA 300

241 GGAAGTGGTAAAGTCATCTTAGAGAAATATACAGATATGGAATGTATGTGATGA 300

301 GGTAAACAGTCCAAAGAGGATCTAGAGATTTGCAAGTTCAATTTGCCGACAACAT 360

301 GGTAAACAGTCCAAAGAGGATCTAGAGATTTGCAAGTTCAATTTGCCGACAACAT 360

361 AAGTGTGACAAATATCATCTTGGCTAATAACAGGCTGTATTTGGCTCTCTTATTT 420

361 AAGTGTGACAAATATCATCTTGGCTAATAACAGGCTGTATTTGGCTCTCTTATTT 420

421 GAAGGCTTGGGAATGAAACTGATATTGTCACATGCGGGCATATGTGGATCAGATGA 480

421 GAAGGCTTGGGAATGAAACTGATATTGTCACATGCGGGCATATGTGGATCAGATGA 480

481 TGGCATAACTTTGAATTTCCGATGTGATATCAGATGACAAAGTTTGCAGTGTGGGAAG 540

481 TGGCATAACTTTGAATTTCCGATGTGATATCAGATGACAAAGTTTGCAGTGTGGGAAG 540

541 TACAGAAGTTTCAATTTCAATTTGGCATTTGGCATTTGAAGAAATTTTCTTTTATCTAG 600

541 TACAGAAGTTTCAATTTCAATTTGGCATTTGGCATTTGAAGAAATTTTCTTTTATCTAG 600

601 TGGTTACAGTGCATATAAATCTCAGCTTTCATATGAAAAATATATGCGAGTGTGTCTCA 660

601 TGGTTACAGTGCATATAAATCTCAGCTTTCATATGAAAAATATATGCGAGTGTGTCTCA 660

661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTATTTTGGTGTCTCTCGGAT 720

661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTATTTTGGTGTCTCTCGGAT 720

721 CTATAAGAGACTTGAATACTCTCTTATAGCTTCTTAAAGAACTCTCTGATGATCAGTG 780

721 CTATAAGAGACTTGAATACTCTCTTATAGCTTCTTAAAGAACTCTCTGATGATCAGTG 780

781 GGTGAGGACAAACAGATTTCCCTCCATCTTGGATAGGCTATCTCTAGCTTATGTTTGA 840

781 GGTGAGGACAAACAGATTTCCCTCCATCTTGGATAGGCTATCTCTAGCTTATGTTTGA 840

841 GTTCCGTAGGGGTGCTGCTTCCAAATTCGAGGAAAGTTTTCCTCATATGCGAAGCG 900

841 GTTCCGTAGGGGTGCTGCTTCCAAATTCGAGGAAAGTTTTCCTCATATGCGAAGCG 900

901 TGAATAACAATATTACTTTACAGACTGGTTTACCTTTCTCTCTCAAAAACTCGCTCT 960

901 TGAATAACAATATTACTTTACAGACTGGTTTACCTTTCTCTCTCAAAAACTCGCTCT 960

961 GGTTCCTCAATGTCCAGCTCCGGAAGAAATTTCAATTCCTACAGATTTTGTTCAAAAT 1020

961 GGTTCCTCAATGTCCAGCTCCGGAAGAAATTTCAATTCCTACAGATTTTGTTCAAAAT 1020

1021 TAGTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080

1021 TAGTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080

1081 ATTAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140

1081 ATTAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140

1141 CTATATAAAGAGTGTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200

1141 CTATATAAAGAGTGTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200

1201 TCTCAAGGTTAGACAACTCCAAATCTCGGTCCTACATTTAGATGATGATGATGATGATGAT 1260

1201 TCTCAAGGTTAGACAACTCCAAATCTCGGTCCTACATTTAGATGATGATGATGATGATGAT 1260

1261 TGTAAAGAGGTTCTAGTAAACACCATGCAAGTTTATTTTGTGGTCCAGAGGTTAAATGT 1320

1261 TGTAAAGAGGTTCTAGTAAACACCATGCAAGTTTATTTTGTGGTCCAGAGGTTAAATGT 1320

1321 TTCCAATCGGGTTCTCCGCAATTTTCTGAGACATAGTAACTTTCTTCGTTGTTCTTT 1380

1321 TTCCAATCGGGTTCTCCGCAATTTTCTGAGACATAGTAACTTTCTTCGTTGTTCTTT 1380

1381 TGTGTAGTGGAGTGGGAGAACTGATTTCTACAGACTTATACCAAAAGCAAGTACTGG 1440

1381 TGTGTAGTGGAGTGGGAGAACTGATTTCTACAGACTTATACCAAAAGCAAGTACTGG 1440

1441 AAGTGTGTGACGACAAACATCTATGAGAGGATCTTATCACTCTCGGAAAGGCTTGT 1500

1441 AAGTGTGTGACGACAAACATCTATGAGAGGATCTTATCACTCTCGGAAAGGCTTGT 1500

1501 AATTGGTGATAAAAATTTGAATTTCTTGCATTTTCATGAGCAGCTTTCGCGGATAATTC 1560

Db 1501 AATTGGTGATAAAAAATTGAATTCCTTGCAATTTTCATCGAGCCAGTTCGGGGAATAATTC 1560
QY 1561 AGTGTGATGTTTGGCATCAAGACCTGGCCTTACTGCAAAATGATATAAGAGCTTGGATGG 1620
Db 1561 AGTGTGATGTTTGGCATCAAGACCTGGCCTTACTGCAAAATGATATAAGAGCTTGGATGG 1620
QY 1621 TGATTTTTCGCAGATCAAGAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680
Db 1621 TGATTTTTCGCAGATCAAGAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680
QY 1681 TTCTCTCAGAGAGACTTTGAGTGTCTTCTAGGATGAGATTGAAGTATATCCCGATGTAAA 1740
Db 1681 TTCTCTCAGAGAGACTTTGAGTGTCTTCTAGGATGAGATTGAAGTATATCCCGATGTAAA 1740
QY 1741 GGTTCATGAAACAGACTATGCTTTCTGATGGAATGGAATGTAATCTGGTGCATTTGC 1800
Db 1741 GGTTCATGAAACAGACTATGCTTTCTGATGGAATGGAATGTAATCTGGTGCATTTGC 1800
QY 1801 TCATAGAGTTGGCTCAAAATGTCGCTTCAATATACCCCATCTGTTTCCAGATTGCTTA 1860
Db 1801 TCATAGAGTTGGCTCAAAATGTCGCTTCAATATACCCCATCTGTTTCCAGATTGCTTA 1860
QY 1861 TGTGTGATATAAGGTGTTTGGGTGTTGATCCGGATTCAATGAAGTGTCTTTGAG 1920
Db 1861 TGTGTGATATAAGGTGTTTGGGTGTTGATCCGGATTCAATGAAGTGTCTTTGAG 1920
QY 1921 AAAGAGCATGTGGAATATGAATCAGACACATAAAGTTAGATGCTTGGATGGAGCAA 1980
Db 1921 AAAGAGCATGTGGAATATGAATCAGACACATAAAGTTAGATGCTTGGATGGAGCAA 1980
QY 1981 ATATCAGCCTTGTTATCTTAACTGTCACACTGATTAGCTCTTGTCTACACTTGGAGTAA 2040
Db 1981 ATATCAGCCTTGTTATCTTAACTGTCACACTGATTAGCTCTTGTCTACACTTGGAGTAA 2040
QY 2041 AGATGAAGTTCGACACAGCAAGCAAGCAAGCTGTAGATGAGCTTATCTTGA 2100
Db 2041 AGATGAAGTTCGACACAGCAAGCAAGCAAGCTGTAGATGAGCTTATCTTGA 2100
QY 2101 TGAATCTTTGAAGGCACAGAGGCTTTTGAATTTGATGCTCTCGAGAGCAACTAAATAT 2160
Db 2101 TGAATCTTTGAAGGCACAGAGGCTTTTGAATTTGATGCTCTCGAGAGCAACTAAATAT 2160
QY 2161 TCTCAAGGCAATGCTAACTGTGGTTTATGAAGCTGATGCTGAGCCCTTCTTCTCAATGAT 2220
Db 2161 TCTCAAGGCAATGCTAACTGTGGTTTATGAAGCTGATGCTGAGCCCTTCTTCTCAATGAT 2220
QY 2221 GTTGCAAACTTCCGGGCATCAAGTGTCTCGATTTCGGACTAGATCAAGAATATTTAT 2280
Db 2221 GTTGCAAACTTCCGGGCATCAAGTGTCTCGATTTCGGACTAGATCAAGAATATTTAT 2280
QY 2281 TCCAAATGGAAGCAATGATGGGATGTTTGGATGAATCCAGAACCTTGGGAATATGGTCA 2340
Db 2281 TCCAAATGGAAGCAATGATGGGATGTTTGGATGAATCCAGAACCTTGGGAATATGGTCA 2340
QY 2341 GGTGTTTGTTCAGTTTACTGTTGCTGGACATGGAGATTTCGACGATTACATCCCAT 2400
Db 2341 GGTGTTTGTTCAGTTTACTGTTGCTGGACATGGAGATTTCGACGATTACATCCCAT 2400
QY 2401 TAATAACAGCAGATCCACCAACAGTAATTTCAATCTGGAAGGAAATGTGGTTGTCAAA 2460
Db 2401 TAATAACAGCAGATCCACCAACAGTAATTTCAATCTGGAAGGAAATGTGGTTGTCAAA 2460
QY 2461 AAATCCATGCTTCATCTCTGTTGATATCGTGTAAAGGCTGTAAATGTTTCGAGCGCT 2520
Db 2461 AAATCCATGCTTCATCTCTGTTGATATCGTGTAAAGGCTGTAAATGTTTCGAGCGCT 2520
QY 2521 GCACCATGTTGATGTTGTTGTTATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580
Db 2521 GCACCATGTTGATGTTGTTGTTATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580
QY 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGGATCAAGATGAT 2640
Db 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGGATCAAGATGAT 2640

QY 2641 CCGCCCAAGGCAAGTCCAGCCGATGGATATCTCTCAGCAACCCAGCATACAGTTGGACCA 2700
Db 2641 CCGCCCAAGGCAAGTCCAGCCGATGGATATCTCTCAGCAACCCAGCATACAGTTGGACCA 2700
QY 2701 TGATGTCACAATTCAGGAAGTTGAAGTACTTCCCAACTATATGTTGAATGACAGTTT 2760
Db 2701 TGATGTCACAATTCAGGAAGTTGAAGTACTTCCCAACTATATGTTGAATGACAGTTT 2760
QY 2761 GGGATCATAGCAATGCCATGCTGATTTTCAGACAGAGAAGCTGATATGGCCATGAG 2820
Db 2761 GGGATCATAGCAATGCCATGCTGATTTTCAGACAGAGAAGCTGATATGGCCATGAG 2820
QY 2821 TGATCCATGCAAAAACCTTGTGAGCTCTTCAATTGCACTGAGCTTCCAAAGACTGG 2880
Db 2821 TGATCCATGCAAAAACCTTGTGAGCTCTTCAATTGCACTGAGCTTCCAAAGACTGG 2880
QY 2881 TGTTCCTCCTGAAATACCATCTCAGTTGCGCCCTTAAAGAAATACCAGACTTCATGGATAA 2940
Db 2881 TGTTCCTCCTGAAATACCATCTCAGTTGCGCCCTTAAAGAAATACCAGACTTCATGGATAA 2940
QY 2941 GCGGCAAGACACAGCTATATCTCAGAAAGAGTTATTTGGAAGCTTTTCAGGAAGTGA 3000
Db 2941 GCGGCAAGACACAGCTATATCTCAGAAAGAGTTATTTGGAAGCTTTTCAGGAAGTGA 3000
QY 3001 GGACAAAGCACCTCAGGCTAGCTCTATCGGACCTTCAAGAGAGATGTTGCAAGAGATC 3060
Db 3001 GGACAAAGCACCTCAGGCTAGCTCTATCGGACCTTCAAGAGAGATGTTGCAAGAGATC 3060
QY 3061 ATATGATGCTGATATGGAAGTTGATGGAATTTGAAGTTACATTTGCAAGGCTTTTGACTA 3120
Db 3061 ATATGATGCTGATATGGAAGTTGATGGAATTTGAAGTTACATTTGCAAGGCTTTTGACTA 3120
QY 3121 CAAAACTCAATATGACACACAGCTGGTAAATTTAATGGACTACTATGCAATAAACAAGA 3180
Db 3121 CAAAACTCAATATGACACACAGCTGGTAAATTTAATGGACTACTATGCAATAAACAAGA 3180
QY 3181 GCCTGAAATACTTGTGCTGCTATATGAAGGCTATCAAAAACCTTTTGACCCGCAAGAAAGA 3240
Db 3181 GCCTGAAATACTTGTGCTGCTATATGAAGGCTATCAAAAACCTTTTGACCCGCAAGAAAGA 3240
QY 3241 TGCTGAGGCCATTAGTGTGCTGTGAGGCCCTTGAGGAGGAGGCAAGAGCTGGTTCAA 3300
Db 3241 TGCTGAGGCCATTAGTGTGCTGTGAGGCCCTTGAGGAGGAGGCAAGAGCTGGTTCAA 3300
QY 3301 GAGCGTAATGATATAGATGACATGTTTACAAAAGCTTCGGCTTGGTACCACGTTACATA 3360
Db 3301 GAGCGTAATGATATAGATGACATGTTTACAAAAGCTTCGGCTTGGTACCACGTTACATA 3360
QY 3361 TCATCCTACATATGGGGTTGCTACATCAGGGGTTGAAAAGAGCTCATTTTCATTTAGCTT 3420
Db 3361 TCATCCTACATATGGGGTTGCTACATCAGGGGTTGAAAAGAGCTCATTTTCATTTAGCTT 3420
QY 3421 TCCCTGGTGTGTTTATGACACAGCTAAATCCAGATTAAAGAGGACAAAGCCTAACAGGCC 3480
Db 3421 TCCCTGGTGTGTTTATGACACAGCTAAATCCAGATTAAAGAGGACAAAGCCTAACAGGCC 3480
QY 3481 AGTTCTCACTTGTCACTCTCAGGGCTCACTGAGTCACAGATTAGTGTGTAATGAGA 3540
Db 3481 AGTTCTCACTTGTCACTCTCAGGGCTCACTGAGTCACAGATTAGTGTGTAATGAGA 3540
QY 3541 TTCCAGTCGAGCGTTAAGCTGATATATATAATAGTATAGGGTGTGATCATATAAGAAAAC 3600
Db 3541 TTCCAGTCGAGCGTTAAGCTGATATATAATAGTATAGGGTGTGATCATATAAGAAAAC 3600
QY 3601 TGTATGCAATTTGTGACTACCTTTTGTCTTAAACCTGCAATGAAGCTGCAACATATATGC 3660
Db 3601 TGTATGCAATTTGTGACTACCTTTTGTCTTAAACCTGCAATGAAGCTGCAACATATATGC 3660
QY 3661 AGTACTCTFAAGAAAACAGATGTACAGCTAAGTAAATATGATGTGATTTGAGTTTTCATC 3720
Db 3661 AGTACTCTFAAGAAAACAGATGTACAGCTAAGTAAATATGATGTGATTTGAGTTTTCATC 3720

QY 3721 TTCTCTCTAA 3731
|||||
Db 3721 TTCTCTCTAA 3731
|||||

RESULT 3
NTA011576
LOCUS Nicotiana tabacum mRNA for RNA-directed RNA polymerase. linear PLN 07-JAN-1999
DEFINITION
ACCESSION AJ011576
VERSION AJ011576.1 GI:4138281
KEYWORDS drp gene; RNA-directed RNA polymerase.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 3505)
Wassenegger, M.
Direct Submission
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
Martinsried, FRG
2 (bases 1 to 3505)
Schiebel, W., Pelissier, T., Riedel, L., Thalmair, S., Schiebel, R.,
Kempe, D., Lottspeich, F., Saenger, H.L. and Wassenegger, M.
Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from Tomato
Unpublished
Location/Qualifiers
1. .3505
/organism="Nicotiana tabacum"
/cultivar="Petit Havana SRI"
/db_xref="taxon:4097"
/tissue_type="leaves"
20. .3370
/gene="RdRp"
20. .3370
/gene="RdRp"
/codon_start=1
/evidence="experimental"
/product="RNA-directed RNA polymerase"
/protein_id="CAA09697.1"
/db_xref="GI:4138282"
/translation="MGKTIQVGFPGYLLSAEAVKKFLENHTGNTGVYALEVKQSGGR
RAFAQVQFANNSAEFLIDASKGLYVGSYLKAMKTDIVQPTVYEHMDGVTLNF
GGQISDKKFAVLGSDVSKFGIGLKKIYFFLSHASADYKKLQSYENWQVYLRHPYG
NQAFLLQFGAPRIYKRLSDSCYSFFKETPDQDQWRTDFTPSWIGLSSCLLEFR
NGVQLPSESFFYKESNMNFILQTFGTFSSQKALVPIVHLPEGLPYKILFKI
SSLIQHCPCFGLAFNFFQLVPRRRTACIEHALEKLYLKECCYDPVRLWTQYD
EYLKQRLPKSPIITLDDGLVYVRVVTPCVYFCGPEVYNSRLVNSYSEDINFL
RYSFVDEWEKIHSTDLPRAGTNGRTDIYERILSTLRNGFIIGDKRFEFLAFSSS
QLRDSNWMKFSRPGLTANDIRTMGDFQIRNVAKYARLGQSGSRETLVSGRHE
VEIPIVACSLGNTYIFSDGIGKISADFAHRTYKGLQYTPSSQYRIGYKGVYA
VPIYSMKLUSLRKMLKYESNNIKLDVLGWSKYPCYLNRQLVTLTLTGVIDDVLQ
KQNEAVDQDLALHDSLKAQALEMSPGENTNLKMLNCGYNPDAPFFLSMLQTF
RASKLDRTRTRIFTPNGRTWMGCLDESRTLEYGVVFQVFSAGRRQFFFEESHFND
SGSANDGFIKGNVYVAKNCPCLHPGDIRVLRADVPAALHMDVDCVYFPQKGRHPNE
CSGSDIDGDIYFVCDPDLIPROQSMYTPAPTQLDHDVTIEVEEYFTNYIIND
SLGIANAHVADREPDMASDRCKLAQLFSIAVDPEKTPGVPAEIPISQRPKEYPD
FMKPKPTPEKVIKGLQKVKNNKTPQASSIATFTRDVARKSYDSMDIVDGFEDI
DEAFYKSYDNKGLMDYGIKTEAELISGLGIMKASKTFDRRDKDAEAGIYAVACSLR
KEARAFYKRSRGIIDMLAKASAWHYVTYHTYGLYNEGLKRDHFIISFPCVYDOLIO
IRKAKRKPVPHLSSLSGSQLSKLVIV"
BASE COUNT 1020 a 622 c 785 g 1078 t
ORIGIN

Query Match 2.5%; Score 94; DB 8; Length 3505;
Best Local Similarity 100.0%; Pred. No. 2.7e-39;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1516 ATTTCGATTTCTTCATTCAGCCAGCTTCGGGATAAATTCAGTGTGGATGTTTGC 1575
|||||

Db 1342 ATTTCGATTTCTTCATTCAGCCAGCTTCGGGATAAATTCAGTGTGGATGTTTGC 1401
QY 1576 ATCAAGACCTGCCTTACTGCATAATGATATAAGA 1609
|||||
Db 1402 ATCAAGACCTGCCTTACTGCATAATGATATAAGA 1435
|||||

RESULT 4
PHY011979
LOCUS Petunia hybrida RdRp gene, partial. linear PLN 07-JAN-1999
DEFINITION
ACCESSION AJ011979
VERSION AJ011979.1 GI:4138342
KEYWORDS RdRp gene; RNA-directed RNA polymerase.
SOURCE Petunia x hybrida.
ORGANISM Petunia x hybrida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
1 (bases 1 to 4579)
Wassenegger, M.
Direct Submission
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
Martinsried, FRG
2 (bases 1 to 4579)
Schiebel, W., Pelissier, T., Riedel, L., Thalmair, S., Schiebel, R.,
Kempe, D., Lottspeich, F., Saenger, H.L. and Wassenegger, M.
Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from Tomato
Unpublished
Location/Qualifiers
1. .4579
/organism="Petunia x hybrida"
/db_xref="taxon:4102"
/tissue_type="leaves"
1. .4579
/gene="RdRp"
join(<1. .1451, 2420. .2570, 3856. .>4579)
/gene="RdRp"
/codon_start=1
/evidence="experimental"
/product="RNA-directed RNA polymerase"
/protein_id="CAA09896.1"
/db_xref="GI:4138343"
/translation="CIEHALEKLYNKAECQYDVRWLTKQYDEYLRKRLPKSPITL
DDGLMYVRVLVTPCKYVFLGPEVNSRVLNRSYSEDINFLRYSFVDEWEKIFSM
LLPRANNGRTDIYERILSTLRNGFIIGDKRFEFLAFSSQLRDSNWMKFSRPGLT
AYDIRTMGDFQIRNVAKYARLGQSGSRETLVSGRHEIKYIPDVEYHGTMYFS
DNKISADFAHRAVAKKCGLOHTPSAQIRIGGKGVVAVDTPSPMKLSLRKSMKYE
SGIKLDVGLGWSKYPCYLNRQLVTLTLTGKDEVPKQNEAVALDAILHDSLKA
QEAALDMSPGENTNLKMLNCGYLPDAEPFLSMMLQTFRASKLDRTRTRIFIPKG
RSMGCLDESRTLEYGVVFQVFTGAGGRQFVEESHFNDSGSANYFILKGNVYVAKN
PCLHPGDIRVLRADVPAALHMDVDCVYFPQKGRHPNEGSGLDGDIFYVCDPDL
IPRQVQPMYTPAPISQLOHDVTIEVEEYFTNYMDGLGVIANAHVYFADREPNN
AMSDPCLAEALQFSIAVDPTKTPGVPAEIPISQRPKEYPDEMEKPKDTYHSEKVGKL
FREYKDTQASSITFTRDVARKSYDITLEDVDFGDIYDEAFYKSYDNKGLMD
YGIKTEAELISGLGIMKASKTFDRRDKDAEAGIYAVACSLRKEARFWRKRSIDDLAK
ASANYHAYHTPYGTCYNEGLKRDHFIISFPCVYD"
1. .1451
/gene="RdRp"
/number=1
1452. .2419
/gene="RdRp"
/number=1
2420. .2570
/gene="RdRp"
/number=2
2571. .3855
/gene="RdRp"
/number=2
3856. .>4579
/gene="RdRp"

Query Match 0.8%; Score 30; DB 8; Length 906;
 Best Local Similarity 100.0%; Pred. No. 0.00042;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3687 TAAGTACTAATATGATGATTCGATTGAGTTT 3716
 Db 47 TAAGTACTAATATGATGATTCGATTGAGTTT 18

RESULT 8
 LOCUS ARI45905/c 25 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 4 from patent US 6218142.
 ACCESSION ARI45905
 VERSION ARI45905.1 GI:15109094
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 25)
 AUTHORS Wassenecker, M., Riedel, L., Schiebel, W. and Sanger, H.L.
 TITLE Nucleic acid molecules encoding polypeptides having the enzymatic activity of an RNA-directed RNA polymerase (RDRP)
 JOURNAL Patent: US 6218142-A 4 17-APR-2001;
 FEATURES Location/Qualifiers
 1..25
 source /organism="unknown"

BASE COUNT 10 a 4 c 7 g 4 t
 ORIGIN

Query Match 0.7%; Score 25; DB 6; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1838 CCATCTGCTTCCAGATTCGTTATG 1862
 Db 25 CCATCTGCTTCCAGATTCGTTATG 1

RESULT 9
 LOCUS ARI45907/c 25 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 6 from patent US 6218142.
 ACCESSION ARI45907
 VERSION ARI45907.1 GI:15109096
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 25)
 AUTHORS Wassenecker, M., Riedel, L., Schiebel, W. and Sanger, H.L.
 TITLE Nucleic acid molecules encoding polypeptides having the enzymatic activity of an RNA-directed RNA polymerase (RDRP)
 JOURNAL Patent: US 6218142-A 6 17-APR-2001;
 FEATURES Location/Qualifiers
 1..25
 source /organism="unknown"

BASE COUNT 4 a 5 c 10 g 6 t
 ORIGIN

Query Match 0.7%; Score 25; DB 6; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GCGATGGAATATCCCTCCAGCACCC 2683
 Db 25 GCGATGGAATATCCCTCCAGCACCC 1

RESULT 10
 LOCUS ARI45906/c 25 bp DNA linear PAT 08-AUG-2001

DEFINITION Sequence 5 from patent US 6218142.
 ACCESSION ARI45906
 VERSION ARI45906.1 GI:15109095
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 25)
 AUTHORS Wassenecker, M., Riedel, L., Schiebel, W. and Sanger, H.L.
 TITLE Nucleic acid molecules encoding polypeptides having the enzymatic activity of an RNA-directed RNA polymerase (RDRP)
 JOURNAL Patent: US 6218142-A 5 17-APR-2001;
 FEATURES Location/Qualifiers
 1..25
 source /organism="unknown"

BASE COUNT 9 a 9 c 4 g 3 t
 ORIGIN

Query Match 0.6%; Score 24; DB 6; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 TGTGGTGTTGATCCGGATTCATC 1902
 Db 24 TGTGGTGTTGATCCGGATTCATC 1

RESULT 11
 LOCUS AC095524/c 112247 bp DNA linear HTG 10-JUL-2002
 DEFINITION Rattus norvegicus clone CH230-7M23, *** SEQUENCING IN PROGRESS ***
 54 unordered pieces.
 ACCESSION AC095524
 VERSION AC095524.3 GI:21716983
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 112247)
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, Z., Lichtarge, O., Lied, C., Liu, J., Liu, M., Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S., Ogih, M., Okwuon, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Fu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,

Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, R., Telford, B., Thomas, R., Thomas, S.,
 Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 112247)
 Worley, K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 112247)
 Worley, K.C.
 Direct Submission
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 9, 2002 this sequence version replaced gi:17942040.

REFERENCE AUTHORS JOURNAL

COMMENT

----- Genome Center of Medicine
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: SCPM
 Center clone name: CH230-7M23
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 61250 bases at least Q40
 Consensus quality: 65255 bases at least Q30
 Consensus quality: 68415 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 54 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1248: contig of 1248 bp in length
 1249 1348: gap of unknown length
 1349 2739: contig of 1391 bp in length
 2740 2839: gap of unknown length
 2840 4110: contig of 1271 bp in length
 4111 4210: gap of unknown length
 4211 5710: contig of 1500 bp in length
 5711 5810: gap of unknown length
 5811 7102: contig of 1292 bp in length
 7103 7202: gap of unknown length
 7203 8664: contig of 1662 bp in length
 8665 8964: gap of unknown length
 8965 10025: contig of 1061 bp in length
 10026 10125: gap of unknown length
 10126 11842: contig of 1717 bp in length
 11843 11942: gap of unknown length
 11943 13271: contig of 1329 bp in length
 13272 13371: gap of unknown length
 13372 14409: contig of 1038 bp in length
 14410 14509: gap of unknown length
 14510 15797: contig of 1288 bp in length
 15798 15897: gap of unknown length
 15898 16919: contig of 1022 bp in length
 16920 17019: gap of unknown length

17020 18674: contig of 1655 bp in length
 18675 18774: gap of unknown length
 18775 20074: contig of 1300 bp in length
 20075 21780: gap of unknown length
 21781 21880: gap of unknown length
 21881 23609: contig of 1729 bp in length
 23610 23709: gap of unknown length
 23710 25341: contig of 1632 bp in length
 25342 25441: gap of unknown length
 25442 26734: contig of 1293 bp in length
 26735 26834: gap of unknown length
 26835 28172: contig of 1238 bp in length
 28173 28172: gap of unknown length
 28173 29339: contig of 1067 bp in length
 29340 29339: gap of unknown length
 29340 30668: contig of 1329 bp in length
 30669 30768: gap of unknown length
 30769 31788: contig of 1020 bp in length
 31789 31888: gap of unknown length
 31889 33345: contig of 1457 bp in length
 33346 33445: gap of unknown length
 33446 34447: contig of 1002 bp in length
 34448 34547: gap of unknown length
 34548 36750: contig of 2203 bp in length
 36751 38214: contig of 1364 bp in length
 38215 38314: gap of unknown length
 38315 39423: contig of 1109 bp in length
 39424 39523: gap of unknown length
 39524 42257: contig of 2734 bp in length
 42258 42357: gap of unknown length
 42358 43881: contig of 1524 bp in length
 43882 43981: gap of unknown length
 43983 46294: contig of 2313 bp in length
 46295 46394: gap of unknown length
 46395 47752: contig of 1358 bp in length
 47753 47852: gap of unknown length
 47853 49218: contig of 1366 bp in length
 49219 49318: gap of unknown length
 49319 51384: contig of 2066 bp in length
 51385 51484: gap of unknown length
 51485 53269: contig of 1785 bp in length
 53270 53369: gap of unknown length
 53370 55675: contig of 2306 bp in length
 55676 55775: gap of unknown length
 55776 57592: contig of 1817 bp in length
 57593 57692: gap of unknown length
 57693 59528: contig of 1836 bp in length
 59529 59628: gap of unknown length
 59629 61385: contig of 1758 bp in length
 61386 61485: gap of unknown length
 61486 63514: contig of 2028 bp in length
 63515 63614: gap of unknown length
 63615 66113: contig of 2499 bp in length
 66114 66213: gap of unknown length
 66214 68443: contig of 2530 bp in length
 68444 72358: gap of unknown length
 72359 72458: contig of 3515 bp in length
 72459 75507: contig of 3049 bp in length
 75508 75607: gap of unknown length
 75608 77933: contig of 2325 bp in length
 77934 78032: gap of unknown length
 78033 80521: contig of 2489 bp in length
 80522 80621: gap of unknown length
 80622 83316: contig of 2695 bp in length
 83317 83416: gap of unknown length
 83417 85800: contig of 2284 bp in length
 85801 85901: gap of unknown length
 85902 89182: contig of 3381 bp in length
 89183 89281: gap of unknown length
 89282 92337: contig of 3076 bp in length

CDS	92457: gap of unknown length	92457: gap of unknown length	complement(join(144..278,360..491,576..673,814..1435,
	92458: contig of 3179 bp in length	92458: contig of 3179 bp in length	1749..1867,2161..2310,2441..2565,2698..2742,2837..2991,
	95637: gap of unknown length	95637: gap of unknown length	2979..3031,3117..3234,3314..3415,3498..3655,3741..3810,
	99368: contig of 3632 bp in length	99368: contig of 3632 bp in length	3894..3985,4054..4104,4138..4420))
	99468: gap of unknown length	99468: gap of unknown length	/note="similar to Na/H antiporter dbj BAA89487.1"
	99469: contig of 4525 bp in length	99469: contig of 4525 bp in length	/codon_start=1
	103993: contig of 4525 bp in length	103993: contig of 4525 bp in length	/evidence=not_experimental
	104093: gap of unknown length	104093: gap of unknown length	/product="F10B6.1"
	104094: contig of 4286 bp in length	104094: contig of 4286 bp in length	/protein_id="AAF79251.1"
	108379: contig of 4286 bp in length	108379: contig of 4286 bp in length	/db_xref="GI:8778242"
108380: gap of unknown length	108380: gap of unknown length	/translation="MTSIIGAALPYKSPKAIASSYSVAENDSSPDVAVIFAGTSLVL	
108480: contig of 3768 bp in length.	108480: contig of 3768 bp in length.	GTACRYLFGNTRVPYTVLLVIGFISGLICLLIKLYSSVISIKPKYKTKENL	
Query Match	0.6%; Score 23; DB 2; Length 112247;	0.6%; Score 23; DB 2; Length 112247;	KLGHGIRKNGINPDLILLAVFLPVLLFESSFSDMVHOKRCOMGVLLAGPGVLIS
	Best Local Similarity 100.0%; Pred. No. 2.2;	Best Local Similarity 100.0%; Pred. No. 2.2;	FCGLALKLPPYVNDWKTSLILGGLLCATDPVAVALLKELGASKMTLLDGESLIM
	Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	NDGVSVVVQLFFKRWGHNSDWGSIIKFLVQNSFGAVGKGLAFGASFWLKAFFIEN
			TVAQITVLTSASYFAYTAQEWAGVSGILTMILGMFFAFAFARTAFKGSOSHSHHF
			EMAAIYANTLVMISGVIIAIESVLSGQTIYKGVSSQFIILRPDFFILFLDVKGGN
			SWSEFLILLYVQSCRVCVVGVVPLLCRSYGLDKWKSIIILTSVGLRAVLOFLDTLG
			GIVFLTVNGSTTQLLHLRLMDTLFATKRILEVTKEMKTKALTAKENLGDDDEEL
			GSADPTVRIHISIKDLGROVPHGVPAAGSLDPTNMDIRIFLNGISOCVWNSS
			HQCTSVYKVAIDSRLLPVGQAAYWEMLDGRTIQCCTANVLMQSDALDLVSTSS
			LSDMQLEPRVHFNYIKFLQSKIIIPKHLVHLIENLJESACIISAFRAHRIARQQ
CDS	132699 bp DNA linear PLN 28-JUN-2000	132699 bp DNA linear PLN 28-JUN-2000	LHIFLNGNSTASTVSENEVEGEAKQFLEDRVDSFPQVLSVLTQKVTYVNLNHLG
	Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome 1, complete sequence.	Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome 1, complete sequence.	YIKLEKVGYSLEKGEVSHLDVQSDLKLLRHPPLKLPNPVDDLITSNPLLKDRSSF
	AC006917	AC006917	RSIAIGETDA"
	AC006917.6	AC006917.6	join(5650..5801,6081..6121,6230..6819)
	GI:4757662	GI:4757662	/note="unknown protein; similar to EST emb F14394.1"
	HTG.	HTG.	/codon_start=1
	Arabidopsis thaliana.	Arabidopsis thaliana.	/evidence=not_experimental
	Arabidopsis thaliana	Arabidopsis thaliana	/product="F10B6.2"
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	/protein_id="AAF79216.1"
	1 (bases 1 to 132699)	1 (bases 1 to 132699)	/db_xref="GI:8778207"
REFERENCE	Chao,Q., Shinn,P., Dunn,P., Buehler,E., Kahn,S., Kim,C., Walker,M., Williams,S., Altafi,H., Araujo,R., Conn,L., Conway,A.B., Gonzalez,A., Hansen,N.F., Huizar,L., Kremenetskaia,I., Lenz,C., Li,J., Liu,S., Luros,S., Rowley,D., Schwartz,J., Toriumi,M., Vystotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.	Chao,Q., Shinn,P., Dunn,P., Buehler,E., Kahn,S., Kim,C., Walker,M., Williams,S., Altafi,H., Araujo,R., Conn,L., Conway,A.B., Gonzalez,A., Hansen,N.F., Huizar,L., Kremenetskaia,I., Lenz,C., Li,J., Liu,S., Luros,S., Rowley,D., Schwartz,J., Toriumi,M., Vystotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.	/translation="MRTPTILLVGAIFSGAGYVRSADSHRYKEGTVPLIYANKV
	Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome 1	Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome 1	GFPNPSTVRYFDLPFCIPEGVKEKEALGVNLDRLVSAKYKLNFRDEKSEVYIC
	Unpublished	Unpublished	NKKLSREVQFRAKVDKYFQMYDDLLPWGFIGKVDKDKJSDSEPKYFLYKHIO
	2 (bases 1 to 132699)	2 (bases 1 to 132699)	FEILYNDRVIEISARDMPHSLVDLHEDKEVDAAEFYTKVKETEPTPPKREKYSMS
	Ecker,J.R.	Ecker,J.R.	SSLPHLEIHWFSIINSCVVLTLTGFLAILMRVLKDNFMK"
	Direct Submission	Direct Submission	join(6881..7029,7140..7296,7392..7876,7961..8180)
	Submitted (25-FEB-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	Submitted (25-FEB-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	/note="similar to transmembrane 9 superfamily member 2
	3 (bases 1 to 132699)	3 (bases 1 to 132699)	gi 4758874; similar to dbj IAV442489.1, gb AI996021.1,
	Ecker,J.R.	Ecker,J.R.	emb F14395.1"
	Direct Submission	Direct Submission	/codon_start=1
REFERENCE	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	/evidence=not_experimental
	4 (bases 1 to 132699)	4 (bases 1 to 132699)	/product="F10B6.3"
	Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C., Chou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.	Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C., Chou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.	/protein_id="AAF79217.1"
	Direct Submission	Direct Submission	/db_xref="GI:8778208"
	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	/translation="MLKRYAQDEAADDQETGWYIHGDVFFPFTNSLFAASLGSG
	5 (bases 1 to 132699)	5 (bases 1 to 132699)	TOLTLFIIFLMLGVGVFYNNRGALFTALVYIYALTSIGIAGTYGASPYCOLEKSW
	Submitted (25-FEB-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	Submitted (25-FEB-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	VRNLLGFCFLGFLTFCLNATLTYTATAALPGTIVTLVTLVTLVTSPLLVLG
	6 (bases 1 to 132699)	6 (bases 1 to 132699)	GIAGKNSAEQAPCTKYPREIPPLVYRSIAPOMAGELPEISAIYIELYIFAS
	Ecker,J.R.	Ecker,J.R.	VGHRITYISLFIIVILLIITATITVATLYFQLAADHDQHWRSFLCGSGTGLFI
	Direct Submission	Direct Submission	YAYCLYIYIARSDNSMGFMQTSFFFGTACIYGFLLMLGTGVFRAALLFVRHYRSIK
REFERENCE	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA	CE"
	7 (bases 1 to 132699)	7 (bases 1 to 132699)	8842..9714
	Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C., Chou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.	Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C., Chou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.	/note="hypothetical protein"
	Direct Submission	Direct Submission	/codon_start=1
	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA	/evidence=not_experimental
	8 (bases 1 to 132699)	8 (bases 1 to 132699)	/product="F10B6.4"
	Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C., Chou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.	Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C., Chou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.	/protein_id="AAF79218.1"
	Direct Submission	Direct Submission	/db_xref="GI:8778209"
	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA	Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA	/translation="MDLSREVDDFIKETIDHSIGLPIKMDVLQKKLYTAESQRRLRE
	9 (bases 1 to 132699)	9 (bases 1 to 132699)	QYLSRLKEKEKVIDLVREASMAQSLKFEENQKIGSERDLVQCKWKSEK
Ecker,J.R.	Ecker,J.R.	FLYHQDRSLMEFGNETDRAEAREVRELEEVKMSDEIKSRLESEDCDLVSIIL	
Direct Submission	Direct Submission	ASVSKDESISLGRIFLEANSDEKCCETLLRKWDOKPTOKFVSLVSWKRIKEK	
COMMENT	Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA	Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA	ECLLMNLAKAEQVELVSEQNRELDRENKRFLLQCCSAERSHSGNSKFNKRSIKMWSSP
	On May 7, 1999 this sequence version replaced gi:4731042.	On May 7, 1999 this sequence version replaced gi:4731042.	IEKRLELSSQEFID"
	Location/Qualifiers	Location/Qualifiers	11067..11906
	1..132699	1..132699	/note="unknown protein; similar to EST gb AI995472.1"
	/organism="Arabidopsis thaliana"	/organism="Arabidopsis thaliana"	/codon_start=1
	/db_xref="taxon:3702"	/db_xref="taxon:3702"	/evidence=not_experimental
	/chromosome="1"	/chromosome="1"	/product="F10B6.5"
			/protein_id="AAF79219.1"
			/db_xref="GI:8778210"

CDS	complement(join(144..278,360..491,576..673,814..1435,	
	1749..1867,2161..2310,2441..2565,2698..2742,2837..2902,	
CDS	2979..3031,3117..3234,3314..3415,3498..3655,3741..3818,	
	3894..3985,4054..4104,4138..4420))	
CDS	/note="similar to Na/H antiporter dbj BAA89487.1"	
	/codon_start=1	
CDS	/evidence=not_experimental	
	/product="F10B6.1"	
CDS	/protein_id="AAF79251.1"	
	/db_xref="GI:8778242"	
CDS	/translation="MTSIIQAALPKYSPKAIASSYSAEINDSSPDVAIVFAGTSIVL	
	GTACRYLNGTRVPYTVLVIGIFLGSIGLICLLIKLSSVISWIKLEKTKHNL	
CDS	KLGHGIRIWNGINPDLIAVFLPFLFESSFMDVHQIKRCMGVLLAGPVLISL	
	FCGLALIKTFPPNWDKTSLLGLGLGATDPVAVALLKELGASKMTLIDGESILM	
CDS	NDGVSVVOLFPEKVMGNDSIIKPLVONSGAVGIGLAFGLASVFWLKFIFND	
	TVAQITVTLASAFYATACEMAGVSIITVILGMFEFAFAFAFKGDSHSLHFHW	
CDS	EMAAANTILVFMLSGVITAEVLSQTSYKGVSSQFIIIRPDFEILFLPDVKGN	
	SWSFLLIYVQLSCVVGVIPLLCRSYGIDWKRESILLTWSGIRAVLQFLPLTG	
CDS	GIVFLIIVNGSTQLLLHLRLMDITATKRILEYKEMMTAKLAVNGSDDEEL	
	GSADWPTVIRHISLKDGRGVNPDHGVAGSLDPTNIMDIRFNGISQCVNNS	
CDS	HCQTSVPKVAIDSRLLPSGVQAAWEMLDGRTQCTANVLMSVDALDVTSS	
	LSDWRGLPRVHPNTYKFSQKIIIPKLVTHLIVERLESACYISSAFRAHRIARQQ	
CDS	LHIFLGNSTASTVINEVEGEAKQFLEDVRDSTPOVLSVLTQVTHYVNLHNG	
	YIKLKVGLGEGKEYSHLDVYQSDLKLLRHPPLSLKLPNVDDLITSNPLKDRSSF	
CDS	RSLAIGETDA"	
	join(5650..5901,6081..6121,6230..6819)	
CDS	/note="unknown protein; similar to EST emb F14394.1"	
	/codon_start=1	
CDS	/evidence=not_experimental	
	/product="F10B6.2"	
CDS	/protein_id="AAF79216.1"	
	/db_xref="GI:8778207"	
CDS	/translation="MRTPTILLVCAILFSGAGYVRSASDRHYKEGDTVPYANKV	
	GPHNPSERYEDFLPCIEGPKERKEALGEVINGLDRLVSAKIPSEDEKESVYC	
CDS	MLKKEEYQKFAVEKDYFYQMYDDLPINGFIGKVDKIDKSDPEKFIYLRHIQ	
	FELLYNDRVIEISARMDPSVLDTEDKEVDAEFMYTVKRETTETPFPRMEKYSMS	
CDS	SSPLHLEIHWFSIINSQVPLILLTGLATILMRVLKNDFMK"	
	join(6881..7029,7140..7296,7392..7876,7961..8180)	
CDS	/note="similar to transmembrane 9 superfamily member 2	
	gi 4758874; similar to dbj JAV442489.1, gb AI996021.1,	
CDS	emb F14395.1"	
	/codon_start=1	
CDS	/evidence=not_experimental	
	/product="F10B6.3"	
CDS	/protein_id="AAF79217.1"	
	/db_xref="GI:8778208"	
CDS	/translation="MLKRYAQDEAADDQETGKVIHGDVFRFPFTNSIFAASLGSG	
	TQFTTIITIFMALGVGYPPYNGRGLFTALVYIYALTSGIAGYSASFCQLEGKSW	
CDS	VRNLLTGFCGLPFLTFCLNTVAITTAALPFGTIVIVILVILVILPFLYVIG	
	GIAGKNSKAEFQACRTKYPREIPPLPYRSAIPQAMAGLFPFSAIYELIYIFAS	
CDS	VMGHRITYIILFIVFIIILVITAFITVALYFQAAEDHQMWRSLFCGGSTGLFI	
	YAYCLYVYVARSMDSGFMQTSFFFGYMACICYGFFLMGLTGVGFRAALLFVRHYSIK	
CDS	CE"	
	8842..9714	
CDS	/note="hypothetical protein"	
	/codon_start=1	
CDS	/evidence=not_experimental	
	/product="F10B6.4"	
CDS	/protein_id="AAF79218.1"	
	/db_xref="GI:8778209"	
CDS	/translation="MDLSREVDFTKETIDHSLGLPISMVLQKKLYTAESORRLRE	
	QYLSVLRKKEKVIDLVRSSAMNAQSLKKFVENOKLSGERDLVNOCKKWEKEC	
CDS	FLYHODRESIMFEGNETDRARESVRELEBEVEKMSDEIKSRLESEDCDVLDSIL	
	AFSVKDSISIGRIELEANSDEKCCFLLKKNQDLKPTTKGTGFLSVSMYKRIEKEK	
CDS	ECILMNLAKAEQEVLLVSQNRDLNRNKKFLNQCSAERSHGSKNFKRKRKSNMSSP	
	IEKRIELSQEYFD"	
CDS	11067..11906	
	/note="unknown protein; similar to EST gb AI995472.1"	
CDS	/codon_start=1	
	/evidence=not_experimental	
CDS	/product="F10B6.5"	
	/protein_id="AAF79219.1"	
CDS	/db_xref="GI:8778210"	


```
/translation="MDDDFRNWGYEPAATFKNGLGQLQMLSDTRNTKPLPGRDP
NLMGPNQSHQEPPIHMSYNWINOQDKDFKNMLPYTATPNVGNLPSEPSAPSMQ
MNLHLHQTENPFLBEEIYVQKKRKINAKGSTPKAKPRPKDSENNNNNNNT
NVRVKKPAKSKVDLVINGMSQDGLPVICTCTGAPQOCYRCGCGWQSQACCTNIS
MPLPMSKRGARISGRKMSQGAFFKVKLEKLASDGFNGFNPIDLKSHWARHGTNKEV
TIR"
complement(13762..14781)
/sequence="AAAF79248.1"
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F1086.6"
/protein_id="AAF79250.1"
/db_xref="GI:8778241"
/translation="MKLWRAAVLVKDGPSLIAADDLTAAVKATSHDELSIDTESA
QPIYRHVLSPPSSKLVSLVSSRVKTRSPAVALKGLMLMHGFELCKSTVAESIGRL
PDLSSFGESGRNRSKSGFNLFVAFELDRSLFHDGCHNRHNEESVILRLV
IRKQIIVLSLIRKIPIGENMIPVINEAMVYSEIMEIYIGWICRHAEVLPNVIS
KIGTEADLAKIVAKMSQGGELKFKFCFKDGLGVSNAQIPNFPVRIPEADVILHDE
LYRTAMESSEESAEETAEAEIEEIEETKSLDLITLDHNEAPAPPPRVVVDIP
DLISF"
16027..16533
/sequence="AAAF79249.1"
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F1086.7"
/protein_id="AAF79220.1"
/db_xref="GI:8778211"
/translation="MSTCVYRECMRHAALGSYADGCREYSQPSGTGLCVACGCH
PSYHRIDVISPQINHTREPTSLRVKQLARLKWKAETAEERNEEDDEETESTEK
MTYQRRRSKTAQREAMKDYAAKLWGLKALRLEEIRVFCGIVRIYHFTWV
NNKRFYH"
complement(join(17635..17815,17901..18019,18141..18194,
18300..18406,18547..18611,18742..18950))
/sequence="AAAF79249.1"
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F1086.8"
/protein_id="AAF79249.1"
/db_xref="GI:8778240"
/translation="MEKFAFYLWYFQPLQCEPVIWFLFCQMILCKKORILFLFW
YPELHLQOPLTKMVFVWFLRLRWFDKOVCKSSSTKGDLSVSGVARGALGFNE
IVRPVKSGLSESPKSLCSQKSGKVLFGQALKVGTSLTFVGEAVRQKAGNLMGKX
EQLVSESESSFEPMNNKMSQELCVILAKIFGSTAVAIYVGYDYARKVLFPV
WKKDLGSSSCNGSSDEDFDISDF"
complement(join(20071..20409,20499..20573,20648..20721,
20800..20887,20957..21108,21188..21356,21448..21906,
22006..22275,22367..22570,22840..22917))
/sequence="AAAF79248.1"
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F1086.9"
/protein_id="AAF79248.1"
/db_xref="GI:8778239"
/translation="MLEIESPTSLCFTNTTCNALLRELQKIWVDIGESDAEKDRMLM
ELEKCLPIYRKVDEANSKAQLHQSILVSTEARIASLLAALGVNESHSPKAKEGSK
SLKEKLAVRNLEDLRLQKDMKQVLDIAQLEKMSGEISGYSQDLNKTWGSJAL
DEQDLTKLNEYTHLRSLOKESDRNLKVDLYNEVHTLCGVDFGQTVSEVRP
SLHRTHQSNISDDTLGLUHHMHLKTERSVRFQKVDKACNAILLFVDFHFA
LAFPLSVQLKDVAGSLFELNMLDTQOERTKFSYVYVRSESTITPENILSSST
IEQVSAEYDFCNKLSKMLVNRRLTELENCLRAHIEADTSLKSTALDSGT
KRLVDSPELLNIELHNKIEEAHSRKEIDRIDRWLSACEENWLEEXNQDSTRYS
AGRGSHVNLKAEARITVNVKIPSMVDNLKIKTLWEDTRKSLYDGVRLVLEDY
KLTRQOEKREYRDCKMODLLIKRESIYSGKSPRSNSVRKTYNGVGDASVPP
TPRNSAGATNDIMTPRSYSSHRQNGYFKEVRLSTAPLNFVAIKESVSITYTSV
CGSPDPSPLYN"
join(26661..26948,27454..27610,27713..27814,27911..28038,
28342..28470,28565..28632,28747..28969)
/sequence="AAAF79221.1"
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F1086.10"
/protein_id="AAF79221.1"
```

```
/db_xref="GI:8778212"
/translation="MIYIRDTKITTKSTIPFLFELFCFSNLSMAILKHKPVLNVFY
VYNLIIFSHSSTAELRLLQPSKTDGTVSFLVIGDWGRSYNSQVALQMGGEIGE
Query Match
Best Local Similarity 0.6% Score 23; DB 8; Length 132699;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2576 AATGAATGTTCTGGAGTGATTT 2598
|||||
Db 63331 AATGAATGTTCTGGAGTGATTT 63309
|||||
RESULT 13
AC113586/204160 bp DNA linear HTG 26-JUN-2002
LOCUS Mus musculus clone RP23-17D24, WORKING DRAFT SEQUENCE, 12 ordered
pieces.
AC113586
VERSION HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE
1 (bases 1 to 204160)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lechoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE
Submitted (03-NAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL
3 (bases 1 to 204160)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lechoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,C.H., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
```


Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vasiliev, H., Vial, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 26, 2002 this sequence version replaced gi:19881849.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBK

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1378

Center clone name: L7_D_24

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 200742 bases at least Q40

Consensus quality: 202113 bases at least Q30

Consensus quality: 202610 bases at least Q20

Insert size: 203000; agarose-1p

Insert size: 203060; sum-of-contigs

Quality coverage: 7.6 in Q20 bases; agarose-1p

Quality coverage: 7.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 12 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 22703: contig of 22703 bp in length

* 22704 22803: gap of 100 bp

* 22804 23139: contig of 336 bp in length

* 23140 23239: gap of 100 bp

* 23240 24461: contig of 1222 bp in length

* 24462 24561: gap of 100 bp

* 24562 25820: contig of 1259 bp in length

* 25821 25920: gap of 100 bp

* 25921 28586: contig of 2666 bp in length

* 28587 28686: gap of 100 bp

* 28687 33028: contig of 4342 bp in length

* 33029 33128: gap of 100 bp

* 33129 52619: contig of 19491 bp in length

* 52620 52719: gap of 100 bp

* 52720 68399: contig of 15680 bp in length

* 68400 68499: gap of 100 bp

* 68500 95071: contig of 26572 bp in length

* 95072 95171: gap of 100 bp

* 95172 117191: contig of 20200 bp in length

* 117192 117291: gap of 100 bp

* 117292 153491: contig of 38200 bp in length

* 153492 155591: gap of 100 bp

* 155592 204160: contig of 48569 bp in length.

* Location/Qualifiers

* 1. .204160

* /organism="Mus musculus"

* /db_xref="taxon:10090"

* /clone_lib="RP23-17D24"

* /clone_lib="RP23-17D24"

* 1. .22703

* /note="assembly_fragment"

* clone_end:SP6

misc_feature

1. .22703

vector_side:left"
22804..23139
/note="assembly_fragment"
23240..24461
/note="assembly_fragment"
24562..25820
/note="assembly_fragment"
25921..28586
/note="assembly_fragment"
28687..33028
/note="assembly_fragment"
33129..52619
/note="assembly_fragment"
52720..68399
/note="assembly_fragment"
68500..95071
/note="assembly_fragment"
95172..117191
/note="assembly_fragment"
117292..155491
/note="assembly_fragment"
155592..204160
/note="assembly_fragment"
vector_side:right"
clone_end:T7

BASE COUNT 59370 a 40877 c 39993 g 62817 t 1103 others
ORIGIN

Query Match 0.6%; Score 22; DB 2; Length 204160;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 673 TCAAAATGCTCAGTTTCTCCTC 694
|||||

Db 11543 TCAAAATGCTCAGTTTCTCCTC 11522
|||||

RESULT 14

AL732358.3

LOCUS

WPCOMMENT

Sequence split into 4 fragments LOCUS AL732358 Accession AL732358

Fragment Name	Begin	End
AL732358_0	1	110000
AL732358_1	100001	210000
AL732358_2	200001	310000
AL732358_3	300001	354683

Continuation (4 of 4) of AL732358 from base 300001 (AL732358 Homo sapiens chromosome

Query Match 0.6%; Score 21; DB 2; Length 54683;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 576 AGAAATTTTTTTCTTTTAT 596
|||||

Db 33471 AGAAATTTTTTTCTTTTAT 33491
|||||

RESULT 15

AL103999

LOCUS

DEFINITION

AC103999 Homo sapiens chromosome 11 clone RP11-96012 map 11, LOW-PASS

SEQUENCE SAMPLING.

AC103999

AC103999.1 GI:17223287

HTG; HTGS_PHASE0.

SOURCE

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 62512)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 11, clone RP11-96012

Unpublished

FEATURES

source

REFERENCE
AUTHORS

2 (bases 1 to 62512)
 Birren, S., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckhalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Katat, A., Karatas, A., Kells, C., Lacroque, K.,
 Lamas, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
 MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneu, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WITR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L21922

Center Clone name: 96_O_12

* NOTE: This record contains 82 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 673: contig of 673 bp in length
 674 773: gap of 100 bp
 774 1448: contig of 675 bp in length
 1449 1548: gap of 100 bp
 1549 2177: contig of 629 bp in length
 2178 2277: gap of 100 bp
 2278 2837: contig of 610 bp in length
 2838 2987: gap of 100 bp
 2988 3684: contig of 677 bp in length
 3685 3764: gap of 100 bp
 3765 4423: contig of 665 bp in length
 4430 4523: gap of 100 bp
 4530 5195: contig of 666 bp in length
 5196 5295: gap of 100 bp
 5296 5980: contig of 685 bp in length
 5981 6080: gap of 100 bp
 6081 6751: contig of 671 bp in length
 6752 6851: gap of 100 bp
 6852 7512: contig of 661 bp in length
 7513 7612: gap of 100 bp
 7613 8283: contig of 671 bp in length
 8284 8383: gap of 100 bp
 8384 9047: contig of 664 bp in length
 9048 9147: gap of 100 bp

9148 9771: contig of 624 bp in length
 9772 9871: gap of 100 bp
 9872 10518: contig of 647 bp in length
 10519 10618: gap of 100 bp
 10619 11261: contig of 643 bp in length
 11262 11361: gap of 100 bp
 11362 12014: contig of 653 bp in length
 12015 12114: gap of 100 bp
 12115 12792: contig of 678 bp in length
 12793 12892: gap of 100 bp
 12893 13563: contig of 671 bp in length
 13564 13663: gap of 100 bp
 13664 14328: contig of 665 bp in length
 14329 14428: gap of 100 bp
 14429 15098: contig of 670 bp in length
 15099 15198: gap of 100 bp
 15199 15867: contig of 669 bp in length
 15868 15967: gap of 100 bp
 15968 16644: contig of 677 bp in length
 16645 16744: gap of 100 bp
 16745 17393: contig of 664 bp in length
 17394 17493: gap of 100 bp
 17494 18166: contig of 673 bp in length
 18167 18266: gap of 100 bp
 18267 18921: contig of 655 bp in length
 18922 19021: gap of 100 bp
 19022 19693: contig of 672 bp in length
 19694 19793: gap of 100 bp
 19794 20457: contig of 664 bp in length
 20458 20557: gap of 100 bp
 20558 21240: contig of 683 bp in length
 21241 21340: gap of 100 bp
 21341 22011: contig of 671 bp in length
 22012 22111: gap of 100 bp
 22112 22790: contig of 679 bp in length
 22791 22890: gap of 100 bp
 22891 23565: contig of 675 bp in length
 23566 23665: gap of 100 bp
 23666 24352: contig of 687 bp in length
 24353 24452: gap of 100 bp
 24453 25116: contig of 664 bp in length
 25117 25216: gap of 100 bp
 25217 25879: contig of 663 bp in length
 25880 25979: gap of 100 bp
 25980 26623: contig of 644 bp in length
 26624 26723: gap of 100 bp
 26724 27377: contig of 654 bp in length
 27378 27477: gap of 100 bp
 27478 28147: contig of 670 bp in length
 28148 28247: gap of 100 bp
 28248 28906: contig of 659 bp in length
 28907 29006: gap of 100 bp
 29007 29680: contig of 674 bp in length
 29681 29780: gap of 100 bp
 29781 30444: contig of 664 bp in length
 30445 30544: gap of 100 bp
 30545 31204: contig of 660 bp in length
 31205 31304: gap of 100 bp
 31305 31981: contig of 677 bp in length
 31982 32081: gap of 100 bp
 32082 32758: contig of 677 bp in length
 32759 32858: gap of 100 bp
 32859 33518: contig of 660 bp in length
 33519 33618: gap of 100 bp
 33619 34274: contig of 656 bp in length
 34275 34374: gap of 100 bp
 34375 35024: contig of 650 bp in length
 35025 35124: gap of 100 bp
 35125 35783: contig of 659 bp in length
 35784 35883: gap of 100 bp
 35884 36563: contig of 680 bp in length
 36564 36663: gap of 100 bp
 36664 37338: contig of 675 bp in length


```

TITLE
JOURNAL
COMMENT
Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced g1:6910776.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project information
Center project name: L3423
Center clone name: L4_M_16
-----
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 857: contig of 857 bp in length
* 958 957: gap of 100 bp
* 958 1852: contig of 895 bp in length
* 1853 1952: gap of 100 bp
* 1953 2811: contig of 859 bp in length
* 2812 2911: gap of 100 bp
* 2912 3773: contig of 862 bp in length
* 3774 3873: gap of 100 bp
* 3874 4737: contig of 864 bp in length
* 4738 4837: gap of 100 bp
* 4838 5710: contig of 873 bp in length
* 5711 5810: gap of 100 bp
* 5811 6692: contig of 882 bp in length
* 6693 6792: gap of 100 bp
* 6793 7651: contig of 859 bp in length
* 7652 7751: gap of 100 bp
* 7752 8643: contig of 892 bp in length
* 8644 8743: gap of 100 bp
* 8744 9620: contig of 877 bp in length
* 9621 9720: gap of 100 bp
* 9721 10603: contig of 883 bp in length
* 10604 10703: gap of 100 bp
* 10704 11582: contig of 879 bp in length
* 11583 11682: gap of 100 bp
* 11683 12567: contig of 885 bp in length
* 12568 12667: gap of 100 bp
* 12668 13534: contig of 867 bp in length
* 13535 13634: gap of 100 bp
* 13635 14479: contig of 845 bp in length
* 14480 14579: gap of 100 bp
* 14580 15449: contig of 870 bp in length
* 15450 15549: gap of 100 bp
* 15550 16432: contig of 883 bp in length
* 16433 16532: gap of 100 bp
* 16533 17413: contig of 881 bp in length
* 17414 17513: gap of 100 bp
* 17514 18415: contig of 902 bp in length
* 18416 18515: gap of 100 bp
* 18516 19408: contig of 893 bp in length
* 19409 19508: gap of 100 bp
* 19509 20387: contig of 879 bp in length
* 20388 20487: gap of 100 bp
* 20488 21390: contig of 909 bp in length
* 21397 21496: gap of 100 bp
* 21497 22335: contig of 839 bp in length
*
* 22336 22435: gap of 100 bp
* 22436 2295: contig of 860 bp in length
* 23296 23395: gap of 100 bp
* 23396 24296: contig of 901 bp in length
* 24297 24396: gap of 100 bp
* 24397 25267: contig of 871 bp in length
* 25268 25367: gap of 100 bp
* 25368 26243: contig of 876 bp in length
* 26244 26343: gap of 100 bp
* 26344 27175: contig of 832 bp in length
* 27176 27275: gap of 100 bp
* 27276 28150: contig of 875 bp in length
* 28151 28250: gap of 100 bp
* 28251 29158: contig of 908 bp in length
* 29159 29258: gap of 100 bp
* 29259 30151: contig of 893 bp in length
* 30152 30251: gap of 100 bp
* 30252 31148: contig of 897 bp in length
* 31149 31248: gap of 100 bp
* 31249 32137: contig of 889 bp in length
* 32138 32237: gap of 100 bp
* 32238 33086: contig of 849 bp in length
* 33087 33186: gap of 100 bp
* 33187 34115: contig of 929 bp in length
* 34116 34215: gap of 100 bp
* 34216 35090: contig of 875 bp in length
* 35091 35190: gap of 100 bp
* 35191 36057: contig of 867 bp in length
* 36058 36157: gap of 100 bp
* 36158 37002: contig of 845 bp in length
* 37003 37102: gap of 100 bp
* 37103 37953: contig of 851 bp in length
* 37954 38053: gap of 100 bp
* 38054 38945: contig of 892 bp in length
* 38946 39045: gap of 100 bp
* 39046 39946: contig of 901 bp in length
* 39947 40046: gap of 100 bp
* 40047 40936: contig of 890 bp in length
* 40937 41036: gap of 100 bp
* 41037 41905: contig of 869 bp in length
* 41906 42005: gap of 100 bp
* 42006 42906: contig of 901 bp in length
* 42907 43006: gap of 100 bp
* 43007 43885: contig of 879 bp in length
* 43886 43985: gap of 100 bp
* 43986 44824: contig of 839 bp in length
* 44825 44924: gap of 100 bp
* 44925 45812: contig of 888 bp in length
* 45813 45912: gap of 100 bp
* 45913 46787: contig of 875 bp in length
* 46788 46887: gap of 100 bp
* 46888 47800: contig of 913 bp in length
* 47801 47900: gap of 100 bp
* 47901 48795: contig of 895 bp in length
* 48796 48895: gap of 100 bp
* 48896 49767: contig of 872 bp in length
* 49768 49867: gap of 100 bp
* 49868 50813: contig of 946 bp in length
* 50814 50913: gap of 100 bp
* 50914 51818: contig of 905 bp in length
* 51819 51918: gap of 100 bp
* 51919 52792: contig of 874 bp in length
* 52793 52892: gap of 100 bp
* 52893 53779: contig of 887 bp in length
* 53780 53879: gap of 100 bp
* 53880 54764: contig of 885 bp in length
* 54765 54864: gap of 100 bp
* 54865 55785: contig of 921 bp in length
* 55786 55885: gap of 100 bp
* 55886 56772: contig of 887 bp in length
* 56773 56872: gap of 100 bp
* 56873 57820: contig of 948 bp in length
* 57821 57920: gap of 100 bp

```

```

* 57921 58797: contig of 877 bp in length
* 58798 58897: gap of 100 bp
* 58898 59751: contig of 854 bp in length
* 59752 59851: gap of 100 bp
* 59852 60731: contig of 880 bp in length
* 60732 60831: gap of 100 bp
* 60832 61723: contig of 892 bp in length
* 61724 61823: gap of 100 bp
* 61824 62748: contig of 925 bp in length
* 62749 62848: gap of 100 bp
* 62849 63738: contig of 890 bp in length
* 63739 63838: gap of 100 bp
* 63839 64717: contig of 879 bp in length
* 64718 64817: gap of 100 bp
* 64818 65674: contig of 857 bp in length
* 65675 65774: gap of 100 bp
* 65775 66735: contig of 961 bp in length
* 66736 66835: gap of 100 bp
* 66836 67703: contig of 868 bp in length
* 67704 67803: gap of 100 bp
* 67804 68689: contig of 886 bp in length
* 68690 68789: gap of 100 bp
* 68790 69656: contig of 867 bp in length
* 69657 69756: gap of 100 bp

Query Match          0.6%; Score 21; DB 2; Length 86478;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
|||||
Db 75142 AATTTTTCCTTTTATCTA 75162

RESULT 18
AP004079
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) chromosome 2 clone
OJ1067_B01, *** SEQUENCING IN PROGRESS ***
ACCESSION
AP004079.1 GI:15281358
VERSION
HTG; HGVS_PHASE2.
KEYWORDS
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:OJ1067_B01.
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
clone:OJ1067_B01
Published Only in Database (2001)
JOURNAL
AP004079
AUTHORS
Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE
Direct Submission
Submitted (22-AUG-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nars.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7458)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and Rgp-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1..101706
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="OJ1067_B01"

BASE COUNT 27699 a 22928 c 22675 g 28253 t 151 others
ORIGIN

Query Match          0.6%; Score 21; DB 2; Length 101706;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
|||||
Db 29464 AATTTTTCCTTTTATCTA 29484

RESULT 19
AC099102
LOCUS
DEFINITION
Rattus norvegicus clone CH230-183A24, *** SEQUENCING IN PROGRESS
AC099102
Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 122685)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.B., Amarantunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.K., David,R.,
Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Moinet,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,I.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sutton,Gren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usman,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
Worley, K.C.
3 (bases 1 to 122685)
Worley, K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:20976259.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGPI
Center clone name: CH230-183A24
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 63780 bases at least Q40
Consensus quality: 67858 bases at least Q30
Consensus quality: 71587 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
1
1183: contig of 1183 bp in length
1184: gap of unknown length
1283: gap of unknown length
1284: contig of 1139 bp in length
2422: contig of 1139 bp in length
2423: gap of unknown length
2522: gap of unknown length
3976: contig of 1454 bp in length
3977: gap of unknown length
4077: gap of unknown length
5136: contig of 1060 bp in length
5237: gap of unknown length
5237: gap of unknown length
6311: contig of 1075 bp in length
6312: gap of unknown length
6412: contig of 1070 bp in length
7481: contig of 1070 bp in length
7482: gap of unknown length
7581: gap of unknown length
7866: contig of 1185 bp in length
8666: gap of unknown length
8667: gap of unknown length
8667: gap of unknown length
10194: contig of 1328 bp in length
10195: gap of unknown length
10294: gap of unknown length
10295: contig of 1713 bp in length
12008: gap of unknown length
12107: gap of unknown length
13316: contig of 1209 bp in length
13317: gap of unknown length
13416: gap of unknown length
14500: contig of 1084 bp in length
14501: gap of unknown length
14600: gap of unknown length
14601: contig of 1281 bp in length
15882: gap of unknown length
15882: gap of unknown length
17114: contig of 1733 bp in length
17115: gap of unknown length
17814: gap of unknown length
17815: gap of unknown length
19609: contig of 1794 bp in length
19708: gap of unknown length
21280: contig of 1572 bp in length
21281: gap of unknown length
21381: gap of unknown length
23129: contig of 1749 bp in length
23229: gap of unknown length
23130: gap of unknown length
23229: gap of unknown length
23230: gap of unknown length
24676: contig of 1447 bp in length
24677: gap of unknown length
24677: gap of unknown length
24777: contig of 1987 bp in length
26764: gap of unknown length
26854: contig of 1652 bp in length
28515: gap of unknown length
28516: gap of unknown length
30279: contig of 1664 bp in length
30280: gap of unknown length
30379: gap of unknown length
31695: contig of 1316 bp in length
31696: gap of unknown length
31795: contig of 1378 bp in length
33174: gap of unknown length
33273: contig of 1850 bp in length
35123: contig of 1850 bp in length
35124: gap of unknown length
35224: contig of 1564 bp in length
36887: gap of unknown length
36888: gap of unknown length
38463: gap of unknown length
38563: contig of 2499 bp in length
41062: gap of unknown length
41161: gap of unknown length
42908: contig of 1747 bp in length
43008: gap of unknown length
44903: contig of 1895 bp in length
45003: gap of unknown length
45004: gap of unknown length
46419: contig of 1416 bp in length
46519: gap of unknown length
46520: gap of unknown length
48624: contig of 2105 bp in length
48724: gap of unknown length
48725: gap of unknown length
49937: contig of 1272 bp in length
50096: gap of unknown length
52648: contig of 2552 bp in length
52748: gap of unknown length
54799: contig of 2051 bp in length
54800: gap of unknown length
56764: contig of 1865 bp in length
56864: gap of unknown length
60043: contig of 3179 bp in length
60143: gap of unknown length
62303: contig of 2160 bp in length
62403: gap of unknown length
64224: contig of 1821 bp in length
64324: gap of unknown length
65853: contig of 1529 bp in length
65854: gap of unknown length
65954: gap of unknown length
69091: contig of 3138 bp in length
69191: gap of unknown length
71000: contig of 1809 bp in length
71100: gap of unknown length
73605: contig of 2505 bp in length
73705: gap of unknown length
76941: contig of 3236 bp in length
77041: gap of unknown length
81869: contig of 4828 bp in length
81870: gap of unknown length
81970: contig of 2753 bp in length
84822: gap of unknown length
84823: gap of unknown length
87601: contig of 2779 bp in length
87701: gap of unknown length
90207: contig of 2506 bp in length
90307: gap of unknown length
94023: contig of 3716 bp in length
94123: gap of unknown length
97520: contig of 3397 bp in length
97521: gap of unknown length
97620: gap of unknown length
101448: contig of 3828 bp in length
101449: gap of unknown length
101548: gap of unknown length
105682: contig of 4144 bp in length
105693: gap of unknown length
110333: contig of 4541 bp in length
110334: gap of unknown length
113636: contig of 3203 bp in length
113736: gap of unknown length
117149: contig of 3413 bp in length
117149: gap of unknown length

```

```

* 117150 117249: gap of unknown length
* 117250 122685: contig of 5436 bp in length.

Query Match      0.6%; Score 21; DB 2; Length 122685;
Best Local Similarity 100.04; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 GAAGAAATTTTCTTTTCTTTT 594
|||||
Db 1935 GAAGAAATTTTCTTTTCTTTT 1955

RESULT 20
HS960017/c
LOCUS
DEFINITION
  Human DNA sequence from clone 960017 on chromosome Xp11.21-11.22
  Contains EST, CA repeat(DX9991), STS, GSS, complete sequence.
ACCESSION
  AL022166
VERSION
  AL022166.1 GI:4034471
KEYWORDS
  HTG; repeat polymorphism.
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 129226)
  Pavitt.R.
REFERENCE
  Direct Submission
  Submitted (27-JAN-1999) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  requests: clonerequests@sanger.ac.uk
  On Dec 19, 1998 this sequence version replaced gi:4007559.
  During sequence assembly data is compared from overlapping clones.
  Where differences are found these are annotated as variations
  together with a note of the overlapping clone name. Note that the
  variation annotation may not be found in the sequence submission
  corresponding to the overlapping clone, as we submit sequences with
  only a small overlap as described above.
  This sequence is the entire insert of clone 960017. The true right
  end of clone 47K8 is at 100 in this sequence. This sequence has
  been finished according to sequence map criteria as follows. An
  attempt is made to resolve all sequencing problems, such as
  compressions and repeats, but not necessarily within known
  annotated human repeat sequence elements (e.g. Alu). Where the
  feature is ambiguous, there is an annotation using the 'unsure'
  feature key.
  This sequence was generated from part of bacterial clone contigs of
  human chromosome X, constructed by the Sanger Centre Chromosome X
  Mapping Group. Further information can be found at
  http://www.sanger.ac.uk/HGP/ChrX
  960017 is from the library RPC15 constructed at the Roswell Park
  Cancer Institute by the group of Pieter de Jong. For further
  details see http://bacpac.med.buffalo.edu/VECTOR:pcvPAC2.
FEATURES
  Location/Qualifiers
    1..129226
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="X"
      /map="p11.21-11.22"
      /clone="RP5-960017"
      /clone_lib="RPC1-5"
      1..3727
        /note="L1PA6 repeat: matches 2414. .6143 of consensus"
      repeat_region
      3749..3926
        /note="HAL1 repeat: matches 343. .524 of consensus"
      repeat_region
      3796..3994
        /note="L1M4c repeat: matches 1497. .1709 of consensus"
      repeat_region
      complement(4521..4899)
        /note="match: GSS A024874"
      misc_feature
      4609..4853
        /note="L1ME3 repeat: matches 5910. .6160 of consensus"
      repeat_region
      4933..5062
        /note="HAL1 repeat: matches 685. .820 of consensus"
      repeat_region
      5491..5783

```

```

/note="AluJb repeat: matches 1. .293 of consensus"
5491..5581
/note="Single clone region"
5963..6128
/note="L1M4 repeat: matches 2340. .2530 of consensus"
6199..6397
/note="L1M4 repeat: matches 4632. .4843 of consensus"
6505..6616
/note="L1M4 repeat: matches 2744. .2851 of consensus"
6617..6877
/note="Alusq repeat: matches 39. .298 of consensus"
6878..7031
/note="L1M4 repeat: matches 2851. .3019 of consensus"
7033..7459
/note="L1M4 repeat: matches 4387. .4834 of consensus"
7583..7888
/note="AluJo repeat: matches 1. .298 of consensus"
7890..7934
/note="Alu repeat: matches 253. .297 of consensus"
7933..8030
/note="L2 repeat: matches 2658. .2748 of consensus"
8084..8359
/note="AluJb repeat: matches 1. .288 of consensus"
8558..8776
/note="L1ME repeat: matches 5263. .5479 of consensus"
9202..9215
/note="L1M4 repeat: matches 5423. .5437 of consensus"
9382..9392
/note="L1M4 repeat: matches 5437. .5587 of consensus"
9623..9927
/note="Alusq repeat: matches 1. .300 of consensus"
10984..11279
/note="Alusx repeat: matches 1. .295 of consensus"
13148..13430
/note="L1 repeat: matches 4789. .5073 of consensus"
13448..14011
/note="L1MAL0 repeat: matches 5765. .6319 of consensus"
14021..14527
/note="L1MA9 repeat: matches 5789. .6284 of consensus"
14528..14863
/note="MER7A repeat: matches 1. .345 of consensus"
14864..15772
/note="L1MA9 repeat: matches 4811. .5789 of consensus"
15774..16392
/note="L1 repeat: matches 3938. .4556 of consensus"
16393..16679
/note="L1PA8 repeat: matches 5886. .6163 of consensus"
16680..17295
/note="L1 repeat: matches 3323. .3938 of consensus"
17298..18190
/note="L1PA2 repeat: matches 5254. .6146 of consensus"
18218..18855
/note="L1M1 repeat: matches 2659. .3329 of consensus"
18856..19207
/note="L1MAL1 repeat: matches 1. .365 of consensus"
19208..20181
/note="L1M1 repeat: matches 1776. .2659 of consensus"
20313..20532
/note="Alusq/x repeat: matches 1. .303 of consensus"
20533..20619
/note="L1M1 repeat: matches 1832. .1908 of consensus"
20620..20912
/note="Alusx repeat: matches 1. .299 of consensus"
20913..20929
/note="L1M1 repeat: matches 1815. .1832 of consensus"
20922..21026
/note="L1M1 repeat: matches 1377. .1769 of consensus"
21063..21414
/note="L1M1 repeat: matches 1398. .1716 of consensus"
21415..21515
/note="HY3 repeat: matches 1. .101 of consensus"
21516..21851
/note="L1M1 repeat: matches 1051. .1398 of consensus"

```

```

repeat_region 21855..22646
/note="LIPB3 repeat: matches 5358..6143 of consensus"
repeat_region 22647..22691
/note="MER44C repeat: matches 92..139 of consensus"
repeat_region 22692..23830
/note="LIPB3 repeat: matches 4162..5311 of consensus"
repeat_region 23821..24600
/note="LIM4C repeat: matches 1028..1815 of consensus"
repeat_region 24601..24890
/note="LIM4C repeat: matches 5995..6291 of consensus"
repeat_region 24891..24969
/note="LIM4C repeat: matches 952..1028 of consensus"
repeat_region 25001..25086
/note="LIMB6 repeat: matches 5860..5954 of consensus"
repeat_region 25094..26572
/note="LIPB1 repeat: matches 4687..5161 of consensus"
repeat_region 26560..27895
/note="LIMB6 repeat: matches 4549..5881 of consensus"
repeat_region 27921..28396
/note="LIMB8 repeat: matches 5764..6288 of consensus"
repeat_region 28397..30161
/note="LIM4 repeat: matches 2740..4562 of consensus"
repeat_region 30159..31561
/note="LIM4C repeat: matches 1044..2037 of consensus"
repeat_region 31575..35678
/note="LIPB13 repeat: matches 2494..6151 of consensus"
repeat_region 35679..35990
/note="LIM4C repeat: matches 1..312 of consensus"
repeat_region 35991..38399
/note="LIPB13 repeat: matches 10..2494 of consensus"
repeat_region 38408..38560
/note="LIM4C repeat: matches 912..1066 of consensus"
repeat_region 38550..39328
/note="LIM4C repeat: matches 5..771 of consensus"
repeat_region 39793..39954
/note="MER6A repeat: matches 10..189 of consensus"
repeat_region 41246..41985
/note="LIMB8 repeat: matches 5542..6288 of consensus"
repeat_region 42046..42183
/note="MLTIF repeat: matches 403..541 of consensus"
repeat_region 42188..42321
/note="LIM4C repeat: matches 1..137 of consensus"
repeat_region 42413..42583
/note="MLTIF repeat: matches 268..431 of consensus"
repeat_region 42550..42850
/note="MLTIF repeat: matches 1..293 of consensus"
repeat_region 44044..44105
/note="31 copies 2 mer ac 68% conserved"
repeat_region 44195..44473
/note="MER69 repeat: matches 2195..2500 of consensus"
repeat_region 44503..45242
/note="MER69 repeat: matches 1..835 of consensus"
repeat_region 45528..50206
/note="LIMB8 repeat: matches 1468..6146 of consensus"
misc_feature 50628..50835
/note="match: STS L41742"
misc_feature complement(51196..551590)
/note="match: GSS AQ020670 clone R-58112"
repeat_region 52095..52406
/note="THEIC repeat: matches 1..371 of consensus"
repeat_region 53658..53962
/note="MLTIF repeat: matches 96..402 of consensus"
misc_feature complement(53950..54193)
/note="match: STS G03867"
repeat_region 54275..54513
/note="MLTIF repeat: matches 4..249 of consensus"
repeat_region 54610..55080
/note="LIMB2 repeat: matches 5914..6340 of consensus"

```

Query Match 0.6; Score 21; DB 9; Length 129226;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 576 AGAAATTTTCTTTCTTTTAT 596
|||||
Db 18223 AGAAATTTTCTTTCTTTTAT 18203

RESULT 21
AC126008/c
LOCUS 131310 bp DNA linear HTG 21-AUG-2002
DEFINITION Medicago truncatula clone mth2-13e5, WORKING DRAFT SEQUENCE, 19
unordered pieces.
AC126008
VERSION AC126008.9 GI:22380755
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Euphorbiaceae; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE 1 (bases 1 to 131310)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Medicago truncatula BAC Clone mth2-13e5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 131310)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 131310)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 21, 2002 this sequence version replaced gi:22325003.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2042: contig of 2042 bp in length
* 2043 2142: gap of unknown length
* 2143 4623: contig of 2481 bp in length
* 4624 4723: gap of unknown length
* 4724 6733: contig of 2010 bp in length
* 6734 6833: gap of unknown length
* 6834 8852: contig of 2019 bp in length
* 8853 11256: gap of unknown length
* 8953 11257: contig of 2304 bp in length
* 11257 11357: gap of unknown length
* 11357 14040: contig of 2684 bp in length
* 14041 14140: gap of unknown length
* 14141 16181: contig of 2041 bp in length
* 16182 16282: gap of unknown length
* 16283 18707: contig of 2425 bp in length
* 18708 18806: gap of unknown length
* 18807 22430: contig of 3624 bp in length
* 22431 22530: gap of unknown length
* 22531 28441: contig of 5911 bp in length
* 28442 28541: gap of unknown length
* 28542 34988: contig of 6447 bp in length

```



```

* 34989 35088: gap of unknown length
* 35089 42615: contig of 7527 bp in length
* 42616 42715: gap of unknown length
* 42716 51635: contig of 8920 bp in length
* 51636 51735: gap of unknown length
* 51736 5975: contig of 8240 bp in length
* 5976 60075: gap of unknown length
* 60076 66561: contig of 6486 bp in length
* 66562 80547: gap of unknown length
* 80548 80647: contig of 13886 bp in length
* 80648 98223: contig of 17576 bp in length
* 98224 98323: gap of unknown length
* 98324 114237: contig of 15914 bp in length
* 114238 114337: gap of unknown length
* 114338 131310: contig of 16973 bp in length.
FEATURES
    Location/Qualifiers
        1..131310
            /organism="Medicago truncatula"
            /db_xref="taxon:3880"
            /clone="mth2-13e5"
            /clone_lib="Medicago truncatula BAC library H2"
BASE COUNT 43027 a 22119 c 22734 g 41572 t 1858 others
ORIGIN
Query Match      0.6%; Score 21; DB 2; Length 131310;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 TGCATAACTTCAGGGGTATT 143
|||||
Db 29192 TGCATAACTTCAGGGGTATT 29172

RESULT 22
AC117166/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-387J16, *** SEQUENCING IN PROGRESS
AC117166
AC117166.2 GI:21746866
HTG; HGSC_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 135458)
Muzny,B.M., Adams,C., Ali-Oudula,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayelle,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,F., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gyllstrom,K.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Luchterghe,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maneshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,

```

```

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogdu,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 135458)
Worley,K.C.
Direct Submission
Submitted (08-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 135458)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20069678.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVDH
Center clone name: CH230-387J16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 108275 bases at least Q40
Consensus quality: 113987 bases at least Q30
Consensus quality: 117413 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 27 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1652: contig of 1652 bp in length
1653: gap of unknown length
1753: contig of 1142 bp in length
2894: gap of unknown length
2895: gap of unknown length
2995: contig of 1057 bp in length
4031: contig of 1057 bp in length
4151: gap of unknown length
4152: contig of 1891 bp in length
6043: gap of unknown length
6143: contig of 1277 bp in length
7419: gap of unknown length
7519: contig of 1278 bp in length
7820: gap of unknown length
8798: gap of unknown length
8898: contig of 2570 bp in length
11467: gap of unknown length
11567: contig of 2656 bp in length
14223: gap of unknown length
14324: contig of 1867 bp in length
14324: contig of 1867 bp in length

```



```

/note="18 copies 6 mer atat 67% conserved"
11464..11569
/note="53 copies 2 mer ta 68% conserved"
11576..11703
/note="64 copies 2 mer at 71% conserved"
12246..12742
/note="match: GSS: Em:AQ429690"
12466..12761
/note="AluJb repeat: matches 1..298 of consensus"
13264..13725
/note="match: GSS: Em:AQ412829"
complement(13345..13537)
/note="match: GSS: Em:AQ342536"
13405..13443
/note="HSMAR1 repeat: matches 1253..1287 of consensus"
13474..13501
/note="7 copies 4 mer attc 92% conserved"
13844..13873
/note="15 copies 2 mer ca 90% conserved"
13878..14299
/note="L2 repeat: matches 1578..2012 of consensus"
14074..14548
/note="match: GSS: Em:AQ808563"
14470..14562
/note="MLR1H repeat: matches 362..447 of consensus"
14888..15000
/note="L2 repeat: matches 2473..2586 of consensus"
15001..15117
/note="L1PA4 repeat: matches 6030..6146 of consensus"
15118..15251
/note="L2 repeat: matches 2586..2733 of consensus"
16315..16551
/note="MIR repeat: matches 12..251 of consensus"
17577..17862
/note="match: GSS: Em:AQ066119"
19153..19302
/note="MIR repeat: matches 92..259 of consensus"
complement(20295..20698)
/note="match: GSS: Em:AQ808901"
20832..21302
/note="match: GSS: Em:AQ528326"
20943..21095
/note="MIR repeat: matches 88..245 of consensus"
22090..22403
/note="AluX repeat: matches 1..312 of consensus"
22676..22785
/note="MIR repeat: matches 148..260 of consensus"
22786..23092
/note="AluX repeat: matches 1..305 of consensus"
23093..23229
/note="MIR repeat: matches 21..148 of consensus"
23832..24499
/note="match: GSS: Em:AQ029286
match: STS: Em:G51371"
24764..25027
/note="L1M2 repeat: matches 5792..6050 of consensus"
25067..25131
/note="L1PA15 repeat: matches 6088..6152 of consensus"
25132..25456
/note="L1PA16 repeat: matches 5701..6152 of consensus"
25449..25539
/note="L1M4 repeat: matches 5474..5563 of consensus"
27676..27882
/note="MLT1H repeat: matches 205..476 of consensus"
28931..29139
/note="MIR repeat: matches 40..262 of consensus"
31511..31542
/note="8 copies 4 mer aatg 90% conserved"
31694..31790
/note="MERSA repeat: matches 88..184 of consensus"
31805..31959
/note="L2 repeat: matches 2548..2708 of consensus"
complement(31985..32329)

```

```

/note="match: GSS: Em:AQ103099"
32447..32919
/note="match: GSS: Em:AQ785577"
32523..33151
/note="match: GSS: Em:AQ540822"
32533..33003
/note="match: GSS: Em:AQ534839"
32964..33060
/note="L2 repeat: matches 2607..2704 of consensus"
33631..33980
/note="MLT1H repeat: matches 140..547 of consensus"
34288..34451
/note="L1PA4 repeat: matches 5981..6146 of consensus"
34765..34955
/note="MIR repeat: matches 1..236 of consensus"
35237..35500
/note="132 copies 2 mer aa 59% conserved"
35251..35418
/note="28 copies 6 mer agagag 64% conserved"
35267..35446
/note="45 copies 4 mer agaa 64% conserved"
36398..36717
/note="L1R16C repeat: matches 50..374 of consensus"
36846..37128
/note="match: GSS: Em:A2243634"
37133..37394
/note="match: GSS: Em:A2243634"
37717..37899
/note="MLT1J repeat: matches 352..535 of consensus"
38110..38224
/note="MLT1H repeat: matches 1..145 of consensus"
38651..38690
/note="20 copies 2 mer gt 100% conserved"
39212..39848
/note="match: GSS: Em:AQ355762"
39250..39315

Query Match 0.8%; Score 21; DB 9; Length 137723;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1941 AATCAGACACATAAAGTTAG 1961
|||||
Db 78002 AATCAGACACATAAAGTTAG 77982

RESULT 24
AP004023/c 138931 bp DNA linear HTG 21-MAR-2002
LOCUS Oryza sativa (japonica cultivar-group) chromosome 2 clone
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone
OJ1126_D09, ** SEQUENCING IN PROGRESS **, in ordered pieces.
ACCESSION AP004023
VERSION AP004023.1 GI:15130685
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:OJ1126_D09.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
clone:OJ1126_D09
JOURNAL Published Only in Database (2001)
REFERENCE
2 (bases 1 to 138931)
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-2001) Takuji Sasaki, National Institute of
Agricultural Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)

```

```

COMMENT      The nucleotide sequence of this BAC clone was generated by
              combining Monsanto and RGP-Japan sequencing data.
              NOTE: It currently consists of 1 contigs. Gaps between the contigs
              are represented as runs of N. The order of the pieces is believed
              to be correct as given, however the sizes of the gaps between them
              are based on estimates that have provided by the submitter. This
              sequence will be replaced by the finished sequence as soon as it is
              available and the accession number will be preserved.
              * NOTE: This is a 'working draft' sequence.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.

FEATURES             location/Qualifiers
    source            1..138931
    /organism="Oryza sativa (japonica cultivar-group)"
    /cultivar="Nipponbare"
    /db_xref="taxon:39947"
    /chromosome="2"
    /clone="OJ1126.D09"
BASE COUNT      39463 a 30336 c 30892 g 37874 t      566 others
ORIGIN
Query Match      0.6%; Score 21; DB 2; Length 138931;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
|||||
Db 12861 AATTTTTCCTTTTATCTA 12841

RESULT 25
AC110620
LOCUS          AC110620 142234 bp DNA linear PRI 05-JUN-2002
DEFINITION    Homo sapiens BAC clone RP11-683I7 from 2, complete sequence.
ACCESSION     AC110620
VERSION       AC110620.3 GI:21326423
KEYWORDS      HTG.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 142234)
              Sulston, J.E. and Waterston, R.
              Toward a complete human genome sequence
              Genome Res. 8 (11), 1097-1108 (1998)
              99063792
              9847074
REFERENCE     2 (bases 1 to 142234)
              Tran, L. and Kozlowicz, A.
              The sequence of Homo sapiens BAC clone RP11-683I7
              Unpublished (2001)
REFERENCE     3 (bases 1 to 142234)
              Waterston, R.H.
              Direct Submission
              Submitted (13-FEB-2002) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              4 (bases 1 to 142234)
              Waterston, R.H.
              Direct Submission
              Submitted (19-FEB-2002) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              5 (bases 1 to 142234)
              Waterston, R.
              Direct Submission
              Submitted (05-JUN-2002) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Jun 5, 2002 this sequence version replaced gi:18702445.
              ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC

```

Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0683I07

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.Chori.org>
 VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-702L18, 2000 bp overlap; the clone sequenced to the right is RP11-341H1, 2000 bp overlap. Actual start of this clone is at base position 107884 of RP11-702L18; actual end is at base position 33545 of RP11-341H1.

Polymorphisms have been identified between AC084367 and AC110620. Data from AC084367 was used to finish this clone, AC110620.

Location/Qualifiers

FEATURES	source
	1..142234
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-683I7"
	/clone_lib="RPCI-11"
	1..56
repeat_region	/rpt_family="Alu"
repeat_region	142..166
repeat_region	/rpt_family="AT-rich"
repeat_region	176..400
repeat_region	/rpt_family="L1"
repeat_region	402..830
repeat_region	/rpt_family="MER1_type"
repeat_region	654..987
repeat_region	/rpt_family="L1"
repeat_region	1043..1198
repeat_region	/rpt_family="L1"
repeat_region	1322..1432
repeat_region	/rpt_family="MER1_type"
repeat_region	1437..1578
repeat_region	/rpt_family="Tc2"
repeat_region	1581..1876
repeat_region	/rpt_family="Alu"
repeat_region	2007..2511
repeat_region	/rpt_family="L1"
repeat_region	2874..2975
repeat_region	/rpt_family="L1"

```

repeat_region 2976..3321 /rpt_family="Alu"
repeat_region 3573..3986 /rpt_family="L1"
repeat_region 3987..4299 /rpt_family="Alu"
repeat_region 4300..4724 /rpt_family="L1"
repeat_region 4725..5018 /rpt_family="Alu"
repeat_region 5019..5214 /rpt_family="L1"
repeat_region 5244..6154 /rpt_family="L1"
repeat_region 6235..6603 /rpt_family="L1"
repeat_region 6609..6785 /rpt_family="ERVL"
repeat_region 6812..7007 /rpt_family="ERVL"
repeat_region 7008..7196 /rpt_family="L1"
repeat_region 7244..7285 /rpt_family="MaLR"
repeat_region 7286..7583 /rpt_family="Alu"
repeat_region 7584..7854 /rpt_family="MaLR"
repeat_region 8885..9164 /rpt_family="L2"
repeat_region 10357..10545 /rpt_family="Alu"
repeat_region 10810..10837 /rpt_family="AT-rich"
repeat_region 10907..10933 /rpt_family="AT-rich"
repeat_region 10934..11245 /rpt_family="Alu"
repeat_region 11981..12615 /rpt_family="ERVL"
repeat_region 15756..15931 /rpt_family="MERL_type"
repeat_region 17805..17830 /rpt_family="AT-rich"
repeat_region 20488..21782 /note="CpG island (%GC=76.8, o/e=0.90, #CpGs=176)"
repeat_region 20593..20621 /rpt_family="GC-rich"
repeat_region 20577..21026 /rpt_family="GC-rich"
repeat_region 21189..21244 /rpt_family="GC-rich"
repeat_region 21376..21471 /rpt_family="G-rich"
repeat_region 22393..22919 /rpt_family="ERVL"
repeat_region 23225..23321 /rpt_family="MIR"
repeat_region 23500..23629 /rpt_family="MERL_type"
repeat_region 23629..23656 /rpt_family="MERL_type"
repeat_region 23657..23949 /rpt_family="Alu"
repeat_region 23950..24016 /rpt_family="MERL_type"
repeat_region 24306..24743 /rpt_family="ERVL"
repeat_region 24762..24834 /rpt_family="MERL_type"
repeat_region 25123..25244 /rpt_family="MIR"
repeat_region 25288..25995

```

```

/rpt_family="MERL_type"
25996..26326
/rpt_family="Alu"
26327..26352
/rpt_family="MERL_type"
26353..26616
/rpt_family="L1"
26528..27528
/rpt_family="L1"
27530..27632
/rpt_family="Alu"
27633..27772
/rpt_family="L1"
28834..28948

Query Match      0.68  Score 21;  DB 9;  Length 142234;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
      |||||
DB 2029 AATTTTTCCTTTTATCTA 2049

RESULT 26
AC108260
LOCUS
DEFINITION AC108260 150055 bp DNA linear HTG 17-JUL-2002
            Rattus norvegicus clone CH230-190K14, *** SEQUENCING IN PROGRESS
            ***, 69 unordered pieces.
ACCESSION AC108260
VERSION AC108260.3 GI:21743959
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.

REFERENCE
1 (bases 1 to 150055)
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraturge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimge,K., Blankenburg,X., Bonnin,D.,
Bouck,J., Bowie,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Roife,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sissot,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

```

```

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
2 (bases 1 to 150055)
Worley,K.C.
Direct Submission
Submitted (27-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 150055)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:18846337.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GM2I
Center clone name: CH230-190K14
----- Summary Statistics
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 80464 bases at least Q40
Consensus quality: 83462 bases at least Q30
Consensus quality: 85753 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1045: contig of 1045 bp in length
* 1046 1145: gap of unknown length
* 1146 2594: contig of 1449 bp in length
* 2595 2894: gap of unknown length
* 2895 3847: contig of 1153 bp in length
* 3848 3947: gap of unknown length
* 3948 5018: contig of 1071 bp in length
* 5019 5118: gap of unknown length
* 5119 6480: contig of 1362 bp in length
* 6481 6580: gap of unknown length
* 6581 7896: contig of 1116 bp in length
* 7897 7996: gap of unknown length
* 7997 9007: contig of 1211 bp in length
* 9008 9107: gap of unknown length
* 9108 10854: contig of 1747 bp in length
* 10855 10954: gap of unknown length
* 10955 11962: contig of 1008 bp in length
* 11963 12662: gap of unknown length
* 12663 13473: contig of 1411 bp in length
* 13474 13573: gap of unknown length
* 13574 15411: contig of 1838 bp in length
* 15412 15511: gap of unknown length
* 15512 16611: contig of 1100 bp in length
* 16612 16711: gap of unknown length
* 16712 17735: contig of 1024 bp in length
* 17736 17835: gap of unknown length
* 17836 19257: contig of 1422 bp in length
* 19258 19357: gap of unknown length

```

```

19358 21141: contig of 1784 bp in length
* 21142 21241: gap of unknown length
* 21242 22719: contig of 1478 bp in length
* 22720 22819: gap of unknown length
* 22820 23882: contig of 1063 bp in length
* 23883 23982: gap of unknown length
* 23983 25414: contig of 1432 bp in length
* 25415 26921: contig of 1407 bp in length
* 26922 27021: gap of unknown length
* 27022 28161: contig of 1140 bp in length
* 28162 28261: gap of unknown length
* 28262 28663: contig of 1602 bp in length
* 28664 29663: gap of unknown length
* 29664 31472: contig of 1509 bp in length
* 31473 31572: gap of unknown length
* 31573 32708: contig of 1136 bp in length
* 32709 32808: gap of unknown length
* 32809 34029: contig of 1221 bp in length
* 34030 34129: gap of unknown length
* 34130 35943: contig of 1814 bp in length
* 35944 36043: gap of unknown length
* 36044 37310: contig of 1267 bp in length
* 37311 37410: gap of unknown length
* 37411 38831: contig of 1421 bp in length
* 38832 38931: gap of unknown length
* 38932 40451: contig of 1520 bp in length
* 40452 40551: gap of unknown length
* 40552 41719: contig of 1168 bp in length
* 41720 41819: gap of unknown length
* 41820 43453: contig of 1634 bp in length
* 43454 43553: gap of unknown length
* 43554 44779: contig of 1226 bp in length
* 44780 44879: gap of unknown length
* 44880 46754: contig of 1875 bp in length
* 46755 46854: gap of unknown length
* 46855 48580: contig of 1726 bp in length
* 48581 48680: gap of unknown length
* 48681 49932: contig of 1252 bp in length
* 49933 50032: gap of unknown length
* 50033 52279: contig of 2247 bp in length
* 52280 53279: gap of unknown length
* 53280 53623: contig of 1144 bp in length
* 53624 55547: contig of 1924 bp in length
* 55548 55647: gap of unknown length
* 55648 56883: contig of 1236 bp in length
* 56884 56983: gap of unknown length
* 56984 58628: contig of 1645 bp in length
* 58629 58728: gap of unknown length
* 58729 61144: contig of 2416 bp in length
* 61145 61244: gap of unknown length
* 61245 62701: contig of 1457 bp in length
* 62702 62801: gap of unknown length
* 62802 64563: contig of 1762 bp in length
* 64564 64663: gap of unknown length
* 64664 65504: contig of 1841 bp in length
* 65505 66604: gap of unknown length
* 66605 68567: contig of 1963 bp in length
* 68568 68667: gap of unknown length
* 68668 70346: contig of 1679 bp in length
* 70347 70446: gap of unknown length
* 70447 71987: contig of 1541 bp in length
* 71988 72087: gap of unknown length
* 72089 73462: contig of 1375 bp in length
* 73463 73562: gap of unknown length
* 73563 75187: contig of 1625 bp in length
* 75188 75287: gap of unknown length
* 75288 77198: contig of 1911 bp in length
* 77199 77298: gap of unknown length
* 77299 79285: contig of 1987 bp in length
* 79286 79385: gap of unknown length
* 79386 81653: contig of 2268 bp in length

```

[illegible]

30732 30831: gap of unknown length
30832 contig of 1675 bp in length
32507 32608: gap of unknown length
32608 contig of 1421 bp in length
34028 34127: gap of unknown length
34127 contig of 1859 bp in length
35987 36086: gap of unknown length
36086 contig of 1117 bp in length
37204 37303: gap of unknown length
37303 contig of 1071 bp in length
38375 38474: gap of unknown length
38474 contig of 1660 bp in length
40134 40234: gap of unknown length
40234 contig of 1113 bp in length
41347 41447: gap of unknown length
41447 contig of 1595 bp in length
43042 43142: gap of unknown length
43142 contig of 1402 bp in length
44644 44744: gap of unknown length
44744 contig of 1512 bp in length
46156 46256: gap of unknown length
46256 contig of 1633 bp in length
47889 47989: gap of unknown length
47989 contig of 2689 bp in length
50678 50778: gap of unknown length
50778 contig of 1519 bp in length
52297 52397: gap of unknown length
52397 contig of 1343 bp in length
53740 53840: gap of unknown length
53840 contig of 2131 bp in length
55971 56071: gap of unknown length
56071 contig of 2613 bp in length
58684 58784: gap of unknown length
58784 contig of 1849 bp in length
60633 60733: gap of unknown length
60733 contig of 2225 bp in length
63038 63138: gap of unknown length
63138 contig of 2903 bp in length
65961 66061: gap of unknown length
66061 contig of 2762 bp in length
68923 69023: gap of unknown length
69023 contig of 1444 bp in length
70367 70467: gap of unknown length
70467 contig of 2235 bp in length
72722 72822: gap of unknown length
72822 contig of 3037 bp in length
75859 75959: gap of unknown length
75959 contig of 2409 bp in length
78368 78468: gap of unknown length
78468 contig of 2340 bp in length
80908 81008: gap of unknown length
81008 contig of 2483 bp in length
83491 83591: gap of unknown length
83591 contig of 2359 bp in length
85950 86050: gap of unknown length
86050 contig of 2371 bp in length
88421 88521: gap of unknown length
88521 contig of 2703 bp in length
91224 91324: gap of unknown length
91324 contig of 2550 bp in length
93775 93875: gap of unknown length
93875 contig of 2608 bp in length
96482 96582: gap of unknown length
96582 contig of 3141 bp in length
99223 99323: gap of unknown length
99323 contig of 3765 bp in length
103588 103688: gap of unknown length
103688 contig of 3089 bp in length
106777 106877: gap of unknown length
106877 contig of 4113 bp in length
110978 111078: gap of unknown length
111078 contig of 3244 bp in length
114334 114434: gap of unknown length

Query Match 0.6%; Score 21; DB 2; Length 153258;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 TCCTACAGATTTTGTTCAA 1017
Db 131524 TCCTACAGATTTTGTTCAA 131544

RESULT 28
AC121213 155859 bp DNA linear HTG 24-JUL-2002
LOCUS Rattus norvegicus clone CH230-302H2, *** SEQUENCING IN PROGRESS
DEFINITION *** 62 unordered pieces.
AC121213
AC121213.2 GI:21909446
VERSION HTG: HTGS_PHASE1.
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 155859)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denna,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbini,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Hernandez,J., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwuo,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shostakov,N., Slason,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155859)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 155859)

AUTHORS
TITLE
JOURNAL
COMMENT

Worley, K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 19, 2002 this sequence version replaced gi:20806243.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GYLF

Center clone name: CH230-302H2

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 105802 bases at least Q40

Consensus quality: 112567 bases at least Q30

Consensus quality: 116059 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 62 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1058: contig of 1058 bp in length

* 1059 1158: gap of unknown length

* 1159 2256: contig of 1098 bp in length

* 2257 2356: gap of unknown length

* 2357 3389: contig of 1033 bp in length

* 3390 3489: gap of unknown length

* 3490 4531: contig of 1042 bp in length

* 4532 4631: gap of unknown length

* 4632 5694: contig of 1063 bp in length

* 5695 5795: gap of unknown length

* 5796 6836: contig of 1042 bp in length

* 6837 6936: gap of unknown length

* 6937 8415: contig of 1479 bp in length

* 8416 8515: gap of unknown length

* 8516 9750: contig of 1235 bp in length

* 9751 9850: gap of unknown length

* 9851 11533: contig of 1683 bp in length

* 11534 11633: gap of unknown length

* 11634 12715: contig of 1082 bp in length

* 12716 12815: gap of unknown length

* 12816 14283: contig of 1468 bp in length

* 14284 14383: gap of unknown length

* 14384 15475: contig of 1092 bp in length

* 15476 15575: gap of unknown length

* 15576 16881: contig of 1306 bp in length

* 16882 16981: gap of unknown length

* 16982 18003: contig of 1022 bp in length

* 18004 18103: gap of unknown length

* 18104 19391: contig of 1288 bp in length

* 19392 19491: gap of unknown length

* 19492 20800: contig of 1309 bp in length

* 20801 20900: gap of unknown length

* 20901 22705: contig of 1805 bp in length

* 22706 22805: gap of unknown length

* 22806 23833: contig of 1028 bp in length

* 23834 23933: gap of unknown length

* 23934 25955: contig of 2022 bp in length

* 25956 26055: gap of unknown length

* 26056 28403: contig of 2348 bp in length

* 28404 28503: gap of unknown length

* 28504 30037: contig of 1594 bp in length

* 30037: contig of 1594 bp in length

* 30098 30197: gap of unknown length

* 30198 32452: contig of 2255 bp in length

* 32453 32552: gap of unknown length

* 32553 33844: contig of 1292 bp in length

* 33845 33944: gap of unknown length

* 33945 36069: contig of 2125 bp in length

* 36070 36169: gap of unknown length

* 36170 37965: contig of 1796 bp in length

* 37966 38065: gap of unknown length

* 38066 40084: contig of 2019 bp in length

* 40085 40184: gap of unknown length

* 40185 41606: contig of 1422 bp in length

* 41607 41706: gap of unknown length

* 41707 43596: contig of 1890 bp in length

* 43597 43696: gap of unknown length

* 43697 45243: contig of 1547 bp in length

* 45244 45343: gap of unknown length

* 45344 46995: contig of 1552 bp in length

* 46996 46995: gap of unknown length

* 46996 48736: contig of 1741 bp in length

* 48737 48836: gap of unknown length

* 48837 50741: contig of 1905 bp in length

* 50742 50841: gap of unknown length

* 50842 52954: contig of 2113 bp in length

* 52955 53054: gap of unknown length

* 53055 55087: contig of 2033 bp in length

* 55088 55187: gap of unknown length

* 55188 56968: contig of 1781 bp in length

* 56969 57068: gap of unknown length

* 57069 59180: contig of 2112 bp in length

* 59181 59280: gap of unknown length

* 59281 61244: contig of 1964 bp in length

* 61245 61344: gap of unknown length

* 61345 63657: contig of 2313 bp in length

* 63658 63757: gap of unknown length

* 63758 66400: contig of 2643 bp in length

* 66401 66500: gap of unknown length

* 66501 68431: contig of 1931 bp in length

* 68432 68531: gap of unknown length

* 68532 72267: contig of 3736 bp in length

* 72268 72367: gap of unknown length

* 72368 75276: contig of 2909 bp in length

* 75277 75376: gap of unknown length

* 75377 77301: contig of 1925 bp in length

* 77302 77401: gap of unknown length

* 77402 80313: contig of 2912 bp in length

* 80314 80413: gap of unknown length

* 80414 82990: contig of 2477 bp in length

* 82991 85335: contig of 2545 bp in length

* 85336 85635: gap of unknown length

* 85636 88017: contig of 2382 bp in length

* 88018 88117: gap of unknown length

* 88118 91242: contig of 3125 bp in length

* 91243 91342: gap of unknown length

* 91343 94858: contig of 3516 bp in length

* 94859 94959: gap of unknown length

* 94959 98762: contig of 3804 bp in length

* 98763 98862: gap of unknown length

* 98863 102013: contig of 3151 bp in length

* 102014 102113: gap of unknown length

* 102114 105285: contig of 3172 bp in length

* 105286 105385: gap of unknown length

* 105386 110193: contig of 4808 bp in length

* 110194 110293: gap of unknown length

* 110294 114524: contig of 4231 bp in length

Query Match 0.6%; Score 21; DB 2; Length 155859;
Best Local Similarity 100.0%; Fred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 581 TTTTTCCTTTTATCTAGT 601

|||||

misc_feature 29383..34427
/note="assembly_fragment"
34528..38012
/note="assembly_fragment"
38113..43120
/note="assembly_fragment"
43221..49302
/note="assembly_fragment"
49403..56105
/note="assembly_fragment"
56206..64386
/note="assembly_fragment"
64487..71512
/note="assembly_fragment"
71613..79328
/note="assembly_fragment"
79429..96099
/note="assembly_fragment"
96200..114101
/note="assembly_fragment"
114202..134807
/note="assembly_fragment"
134908..155904
/note="assembly_fragment"
BASE COUNT 47807 a 29672 c 29375 g 46845 t 2205 others
ORIGIN
Query Match 0.6%; Score 21; DB 2; Length 155904;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2329 GGAATATGGTCAGGTGTTGT 2349
|||||
Db 63859 GGAATATGGTCAGGTGTTGT 63879
RESULT 30
AC114071
LOCUS
DEFINITION Rattus norvegicus clone CH230-14017, *** SEQUENCING IN PROGRESS
AC114071 157759 bp DNA linear HTG 13-JUL-2002
VERSION AC114071.2 GI:21731941
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 157759)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsfi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louliseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

10864 10963: gap of unknown length
 10964 12292: contig of 1329 bp in length
 12293 12392: gap of unknown length
 12393 13409: contig of 1017 bp in length
 13410 13509: gap of unknown length
 13510 15073: contig of 1564 bp in length
 15074 15173: gap of unknown length
 15174 16589: contig of 1416 bp in length
 16590 16689: gap of unknown length
 16690 18332: contig of 1643 bp in length
 18333 18432: gap of unknown length
 18433 19956: contig of 1524 bp in length
 19957 20056: gap of unknown length
 20057 21151: contig of 1095 bp in length
 21152 21251: gap of unknown length
 21252 22251: contig of 1000 bp in length
 22252 22351: gap of unknown length
 22352 23738: contig of 1387 bp in length
 23739 23838: gap of unknown length
 23839 25427: contig of 1589 bp in length
 25428 25527: gap of unknown length
 25528 27387: contig of 1860 bp in length
 27388 27487: gap of unknown length
 27488 28922: contig of 1435 bp in length
 28923 29022: gap of unknown length
 29023 30162: contig of 1140 bp in length
 30163 30262: gap of unknown length
 30263 31771: contig of 1509 bp in length
 31772 31871: gap of unknown length
 31872 34314: contig of 2443 bp in length
 34315 34414: gap of unknown length
 34415 35957: contig of 1543 bp in length
 35958 36057: gap of unknown length
 36058 38356: contig of 2299 bp in length
 38357 38456: gap of unknown length
 38457 39903: contig of 1447 bp in length
 39904 40003: gap of unknown length
 40004 41599: contig of 1696 bp in length
 41600 41799: gap of unknown length
 41800 43000: contig of 1201 bp in length
 43001 43100: gap of unknown length
 43101 44542: contig of 1442 bp in length
 44543 44642: gap of unknown length
 44643 46604: contig of 1962 bp in length
 46605 46704: gap of unknown length
 46705 48602: contig of 1898 bp in length
 48603 48702: gap of unknown length
 48703 51016: contig of 2314 bp in length
 51017 51116: gap of unknown length
 51117 53668: contig of 2552 bp in length
 53669 53768: gap of unknown length
 53769 55640: contig of 1872 bp in length
 55641 55740: gap of unknown length
 55741 58120: contig of 2379 bp in length
 58121 58219: gap of unknown length
 58220 59888: contig of 1669 bp in length
 59889 59988: gap of unknown length
 59989 61976: contig of 1988 bp in length
 61977 62076: gap of unknown length
 62077 64593: contig of 2507 bp in length
 64584 64683: gap of unknown length
 64684 67559: contig of 2876 bp in length
 67560 67659: gap of unknown length
 67660 71558: contig of 3899 bp in length
 71559 71658: gap of unknown length
 71659 73750: contig of 2092 bp in length
 73751 73850: gap of unknown length
 73851 77404: contig of 3554 bp in length
 77405 77505: gap of unknown length
 77506 80262: contig of 2758 bp in length
 80263 80362: gap of unknown length
 80363 82789: contig of 2427 bp in length
 82790 82889: gap of unknown length

* 82890 85869: contig of 2980 bp in length
 * 85870 85969: gap of unknown length
 * 85970 89739: contig of 3770 bp in length
 * 89740 89839: gap of unknown length
 * 89840 92739: contig of 2890 bp in length
 * 92739 92839: gap of unknown length
 * 92840 94946: contig of 2117 bp in length
 * 94947 95046: gap of unknown length
 * 95047 97684: contig of 2638 bp in length
 * 97685 97784: gap of unknown length
 * 97785 100890: contig of 3106 bp in length
 * 100891 100990: gap of unknown length
 * 100991 103985: contig of 2975 bp in length
 * 103986 104065: gap of unknown length
 * 104066 106355: contig of 2290 bp in length
 * 106356 106455: gap of unknown length
 * 106456 111448: contig of 4993 bp in length
 * 111449 111548: gap of unknown length
 * 111549 115697: contig of 4149 bp in length

Query Match 0.6%; Score 21; DB 2; Length 157759;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 581 TTTTTCCTTTTATCTAGT 601
 Db 137270 TTTTTCCTTTTATCTAGT 137290

RESULT 31
 AC034304 158158 bp DNA linear HTG 22-MAY-2000
 LOCUS Homo sapiens chromosome 8 clone RP11-194B7 map 8, WORKING DRAFT
 DEFINITION SEQUENCE, 11 unordered pieces.
 AC034304
 AC034304.2 GI:8018117
 VERSION HTG; HTGS-PHASE1; HTGS-DRAFT.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens
 ORGANISM
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 158158)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 8, clone RP11-194B7
 Unpublished
 2 (bases 1 to 158158)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
 Boguslavsky, L., Boukhgaiter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castler, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Navlor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schaefer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testay, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission

TITLE
 JOURNAL
 COMMENT
 Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 22, 2000 this sequence version replaced gi.7417882.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, K., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (16-OCR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE AUTHORS

4 (bases 1 to 158371)
Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Canarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
Landers, T., Lehoczyk, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Minova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (17-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Feb 17, 2002 this sequence version replaced gi:16152312.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- GenomE Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2947

Center clone name: 7_K_9

FEATURES

Source

Location/Qualifiers

1. .158371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-7K9"
/clone_lib="RP11 Human Male BAC"

repeat_region

1184..1231

repeat_region

/rpt_family="TTTGn"

repeat_region

1607..1627

repeat_region

2552..2376

repeat_region

/rpt_family="MIR3"

repeat_region

2974..2996

repeat_region

/rpt_family="AT-rich"

repeat_region

complement(3290..3497)

repeat_region

/rpt_family="MIR"

repeat_region

complement(3639..3676)

repeat_region

/rpt_family="MIR"

repeat_region

3933..3954

repeat_region

/rpt_family="AT-rich"

repeat_region

7421..7462

repeat_region

/rpt_family="(CA)n"

repeat_region

complement(8362..8482)

repeat_region
8734..8780
/rpt_family="L2"
repeat_region
complement(9127..9470)
/rpt_family="THE1B"
repeat_region
9827..10183
/rpt_family="THE1B"
repeat_region
complement(10775..11105)
/rpt_family="MER1B"
repeat_region
11147..11568
/rpt_family="MSTA"
repeat_region
12908..13215
/rpt_family="ALUy"
repeat_region
15012..15611
/rpt_family="L2"
repeat_region
15826..15847
/rpt_family="AT-rich"
repeat_region
complement(19223..19339)
/rpt_family="L1MA2"
repeat_region
19917..19953
/rpt_family="(CA)n"
repeat_region
21972..21997
/rpt_family="AT-rich"
repeat_region
22094..22189
/rpt_family="MIR"
repeat_region
complement(22190..22502)
/rpt_family="ALUSx"
repeat_region
22503..22527
/rpt_family="MIR"
repeat_region
complement(22546..22677)
/rpt_family="MERSA"
repeat_region
22622..22692
/rpt_family="MERSA"
repeat_region
23815..23842
/rpt_family="(CA)n"
repeat_region
24479..24506
/rpt_family="AT-rich"
repeat_region
complement(24719..24892)
/rpt_family="MER20"
repeat_region
25709..25753
/rpt_family="MIR"
repeat_region
26575..26610
/rpt_family="AT-rich"
repeat_region
27572..27605
/rpt_family="polypyrimidine"
repeat_region
27629..27704
/rpt_family="L2"
repeat_region
complement(27868..28055)
/rpt_family="MERSA"
repeat_region
28606..28953
/rpt_family="L2"
repeat_region
29786..29986
/rpt_family="MIR"
repeat_region
30391..30421
/rpt_family="AT-rich"
repeat_region
31234..31327
/rpt_family="(GAA)n"
repeat_region
31684..31744
/rpt_family="CT-rich"
repeat_region
31849..31990
/rpt_family="MIR3"
repeat_region
32011..32253
/rpt_family="MIR"
repeat_region
complement(33243..33390)
/rpt_family="MIR"
repeat_region
34041..34210
/rpt_family="L1PB3"
repeat_region
34739..34942
/rpt_family="MIR"
repeat_region
complement(38566..38717)
/rpt_family="MERSA"
repeat_region
39098..39251
/rpt_family="MIR"

```
repeat_region 39972..40277
/rpt_family="Alu9g"
repeat_region 41183..43406
/rpt_family="L1p"
repeat_region 43582..43667
/rpt_family="L2"

Query Match 0.6% Score 21; DB 9; Length 158371;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2329 GGAATATGTCAGGTGTTTGT 2349
|||||
Db 100964 GGAATATGTCAGGTGTTTGT 100984
|||||

RESULT 33
AC129367/c
LOCUS AC129367 160083 bp DNA linear HTG 02-AUG-2002
DEFINITION Rattus norvegicus clone CH230-253D15, *** SEQUENCING IN PROGRESS.
ACCESSION AC129367
VERSION AC129367.1 GI:22001448
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 160083)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbakra,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Edwards,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,
Lozadro,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwionu,G.,
Orgunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinscen,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 160083)

AUTHORS
Worley,K.C.
Direct Submission
Submitted (29-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 160083)
Worley,K.C.
Direct Submission
Submitted (02-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTOW
Center clone name: CH230-253D15
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 114920 bases at least Q40
Consensus quality: 120930 bases at least Q30
Consensus quality: 124914 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1072: contig of 1072 bp in length
* 1073 1172: gap of unknown length
* 1173 2556: contig of 1384 bp in length
* 2557 2656: gap of unknown length
* 2657 3986: contig of 1330 bp in length
* 3987 4086: gap of unknown length
* 4087 5599: contig of 1513 bp in length
* 5600 5699: gap of unknown length
* 5700 6744: contig of 1044 bp in length
* 6843: gap of unknown length
* 6844 8313: contig of 1470 bp in length
* 8314 9472: contig of 1059 bp in length
* 9473 9572: gap of unknown length
* 9573 10703: contig of 1131 bp in length
* 10704 10803: gap of unknown length
* 10804 12122: contig of 1319 bp in length
* 12123 13406: contig of 1184 bp in length
* 13407 13506: gap of unknown length
* 13507 14978: contig of 1472 bp in length
* 14979 15078: gap of unknown length
* 15079 17119: contig of 2041 bp in length
* 17120 17219: gap of unknown length
* 17220 19195: contig of 1976 bp in length
* 19196 19295: gap of unknown length
* 19296 21926: contig of 2631 bp in length
* 21927 22026: gap of unknown length
* 22027 23514: contig of 1488 bp in length
* 23515 23614: gap of unknown length
* 23615 24835: contig of 1220 bp in length
* 24836 24934: gap of unknown length
* 24935 26513: contig of 1579 bp in length
* 26514 27809: contig of 1196 bp in length
* 27810 27909: gap of unknown length
```

Support **and** **Person**

REFERENCE
3 (bases 1 to 162600)
Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

TITLE
JOURNAL

Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gi:21206167.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center Project name: TUEK
Center Clone name: CH230-LA15
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 55% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 82066 bases at least Q40
Consensus quality: 88733 bases at least Q30
Consensus quality: 91777 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 77 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1155: contig of 1155 bp in length
* 1156: 1255: gap of unknown length
* 1256: 2356: contig of 1101 bp in length
* 2357: 2456: gap of unknown length
* 2457: 3545: contig of 1089 bp in length
* 3546: 3645: gap of unknown length
* 3646: 4676: contig of 1031 bp in length
* 4677: 4776: gap of unknown length
* 4777: 5930: contig of 1154 bp in length
* 5931: 6030: gap of unknown length
* 6031: 7318: contig of 1288 bp in length
* 7319: 7418: gap of unknown length
* 7419: 8572: contig of 1154 bp in length
* 8573: 8672: gap of unknown length
* 8673: 9903: contig of 1231 bp in length
* 9904: 10003: gap of unknown length
* 10004: 11063: contig of 1060 bp in length
* 11064: 11163: gap of unknown length
* 11164: 12198: contig of 1035 bp in length
* 12199: 12298: gap of unknown length
* 12299: 13847: contig of 1549 bp in length
* 13848: 13947: gap of unknown length
* 13948: 14950: contig of 1003 bp in length
* 14951: 15050: gap of unknown length
* 15051: 16408: contig of 1358 bp in length
* 16409: 16508: gap of unknown length
* 16509: 17676: contig of 1168 bp in length
* 17677: 17776: gap of unknown length
* 17777: 18907: contig of 1031 bp in length
* 18908: 18907: gap of unknown length
* 18908: 20366: contig of 1459 bp in length
* 20367: 20466: gap of unknown length
* 20467: 21549: contig of 1083 bp in length
* 21550: 21649: gap of unknown length
* 21650: 22658: contig of 1009 bp in length
* 22659: 22758: gap of unknown length
* 22759: 24196: contig of 1437 bp in length
* 24196: 24295: gap of unknown length
* 24296: 25537: contig of 1242 bp in length
* 25538: 25637: gap of unknown length

* 25638: 26918: contig of 1281 bp in length
* 26919: 27018: gap of unknown length
* 27019: 28091: contig of 1073 bp in length
* 28092: 28191: gap of unknown length
* 28192: 29567: contig of 1376 bp in length
* 29568: 29667: gap of unknown length
* 29668: 31186: contig of 1519 bp in length
* 31187: 31286: gap of unknown length
* 31287: 32779: contig of 1493 bp in length
* 32780: 32879: gap of unknown length
* 32880: 34510: contig of 1631 bp in length
* 34511: 34610: gap of unknown length
* 34611: 36269: contig of 1659 bp in length
* 36270: 36369: gap of unknown length
* 36370: 37749: contig of 1380 bp in length
* 37750: 37849: gap of unknown length
* 37850: 39396: contig of 1547 bp in length
* 39397: 39496: gap of unknown length
* 39497: 40601: contig of 1105 bp in length
* 40602: 40701: gap of unknown length
* 40702: 42561: contig of 1860 bp in length
* 42562: 42661: gap of unknown length
* 42662: 44434: contig of 1793 bp in length
* 44435: 44534: gap of unknown length
* 44535: 46043: contig of 1489 bp in length
* 46044: 46143: gap of unknown length
* 46144: 47705: contig of 1562 bp in length
* 47706: 47805: gap of unknown length
* 47806: 49580: contig of 1775 bp in length
* 49581: 49680: gap of unknown length
* 49681: 51430: contig of 1770 bp in length
* 51431: 51550: gap of unknown length
* 51551: 53143: contig of 1593 bp in length
* 53144: 53243: gap of unknown length
* 53244: 54806: contig of 1563 bp in length
* 54807: 54906: gap of unknown length
* 54907: 56515: contig of 1609 bp in length
* 56516: 56615: gap of unknown length
* 56616: 58202: contig of 1587 bp in length
* 58203: 58302: gap of unknown length
* 58303: 60165: contig of 1863 bp in length
* 60166: 60265: gap of unknown length
* 60266: 62601: contig of 2336 bp in length
* 62602: 62701: gap of unknown length
* 62702: 64499: contig of 1798 bp in length
* 64500: 64599: gap of unknown length
* 64600: 67052: contig of 2453 bp in length
* 67053: 67152: gap of unknown length
* 67153: 68917: contig of 1765 bp in length
* 68918: 69017: gap of unknown length
* 69019: 71409: contig of 2392 bp in length
* 71410: 71509: gap of unknown length
* 71510: 73977: contig of 2467 bp in length
* 73977: 74077: gap of unknown length
* 74077: 75585: contig of 1509 bp in length
* 75586: 75686: gap of unknown length
* 75687: 77579: contig of 1894 bp in length
* 77580: 77679: gap of unknown length
* 77680: 79357: contig of 1678 bp in length
* 79358: 79457: gap of unknown length
* 79458: 81312: contig of 1855 bp in length
* 81313: 81412: gap of unknown length
* 81413: 83721: contig of 2309 bp in length
* 83722: 83821: gap of unknown length
* 83822: 85452: contig of 1631 bp in length

Query Match 0.6%; Score 21; DB 2; Length 162600;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 TCCCTACAGATTGTTCAA 1017

DB 31768 TCCCTACAGATTGTTCAA 31788

```
RESULT 35
AP001802/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

AP001802
Homo sapiens chromosome 11 clone RP11-100J10 map 11q, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AP001802
AP001802.3 GI:9757489
HTG: HTGS-PHASE1; HTGS-DRAFT.
HOMO sapiens DNA, clone:RP11-100J10.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165699)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 165,699 genomic DNA of 11q
Published Only in Database (2000)
2 (bases 1 to 165699)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (14-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Aug 9, 2000 this sequence version replaced gi:8117473.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gscc.riken.go.jp
----- Project Information
Center project name: HumDrafft11
Center clone name: RP11-100J10
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 162162 bases at least Q40
Consensus quality: 163565 bases at least Q30
Consensus quality: 164180 bases at least Q20
Insert size: 164699; sum-of-contigs
Quality coverage: 11.49x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
11 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 39447 contig of 39447 bp in length
39548 39547: gap of 100 bp
39548 71742: contig of 32195 bp in length
71843 71742: gap of 100 bp
71843 99207: contig of 27365 bp in length
99308 99207: gap of 100 bp
99308 117932: contig of 18625 bp in length
118033 117932: gap of 100 bp
118033 134393: contig of 16361 bp in length
134394 134393: gap of 100 bp
134494 139889: contig of 5396 bp in length
139890 139889: gap of 100 bp
139990 145501: contig of 5512 bp in length
145502 145501: gap of 100 bp
145602 150056: contig of 4455 bp in length
150057 150056: gap of 100 bp
150157 155960: contig of 5804 bp in length
155961 156060: gap of 100 bp
156061 161995: contig of 5935 bp in length
161996 162095: gap of 100 bp
162096 165699: contig of 3604 bp in length.
FEATURES
Location/Qualifiers
source
1..165699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-100J10"
misc_feature
1..39447
/note="assembly_fragment"
39548..71742
/note="assembly_fragment"
71843..99207
/note="assembly_fragment"
99308..117932
/note="assembly_fragment"
118033..134393
/note="assembly_fragment"
134494..139889
/note="assembly_fragment"
139990..145501
/note="assembly_fragment"
145602..150056
/note="assembly_fragment"
150157..155960
/note="assembly_fragment"
156061..161995
/note="assembly_fragment"
162096..165699
/note="assembly_fragment"
BASE COUNT 44435 a 35614 c 36787 g 47863 t 1000 others
ORIGIN
Query Match 0.6%; Score 21; DB 2; Length 165699;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 579 AATTTTTCCTTTTATCTA 599
|||||
Db 108062 AATTTTTCCTTTTATCTA 108042
|||||
RESULT 36
AC095109/c
LOCUS
DEFINITION
AC095109
AC095109.3 GI:21716907
HTG: HTGS-PHASE1.
Norway rat
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Rattus norvegicus clone CH230-7M2, *** SEQUENCING IN PROGRESS ***,
70 unordered pieces.
AC095109
AC095109.3 GI:21716907
HTG: HTGS-PHASE1.
Norway rat
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 168763)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimade,K., Blankenburg,K., Bonnin,D.,
Buck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eantheart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,P., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miser,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogih,M., Okwundu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Roife,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tatney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 168763)
Worley,K.C.
Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 168763)
Worley,K.C.
Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 9, 2002 this sequence version replaced gi:17941975.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCKQ
Center clone name: CH230-7M2
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 107542 bases at least Q40
Consensus quality: 115410 bases at least Q30
Consensus quality: 112372 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 70 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
*
* 1
1194: contig of 1194 bp in length
1294: gap of unknown length
1295: contig of 1079 bp in length
2374: gap of unknown length
2474: contig of 1561 bp in length
4035: gap of unknown length
4135: contig of 1349 bp in length
5483: gap of unknown length
5584: contig of 1430 bp in length
7013: gap of unknown length
7113: contig of 1661 bp in length
8774: gap of unknown length
8775: contig of 1516 bp in length
10390: gap of unknown length
10391: contig of 1852 bp in length
12342: gap of unknown length
12442: contig of 1766 bp in length
14208: gap of unknown length
14309: contig of 1185 bp in length
14309: gap of unknown length
15493: gap of unknown length
15494: contig of 1188 bp in length
1594: gap of unknown length
16881: contig of 1363 bp in length
18244: gap of unknown length
18245: contig of 1447 bp in length
18345: gap of unknown length
19792: contig of 1930 bp in length
21821: gap of unknown length
21921: contig of 1382 bp in length
23303: gap of unknown length
23403: contig of 1472 bp in length
24875: gap of unknown length
24976: contig of 1189 bp in length
26165: gap of unknown length
26265: contig of 2460 bp in length
28244: gap of unknown length
28245: contig of 1718 bp in length
30542: gap of unknown length
30642: contig of 1515 bp in length
32157: gap of unknown length
34514: contig of 2257 bp in length
34614: gap of unknown length
36302: contig of 1688 bp in length
36402: gap of unknown length
38004: contig of 1602 bp in length
38104: gap of unknown length
39979: contig of 1875 bp in length
40079: gap of unknown length
41784: contig of 1705 bp in length
41884: gap of unknown length
43604: contig of 1720 bp in length
43704: gap of unknown length
45685: contig of 1961 bp in length
45765: gap of unknown length
47591: contig of 1826 bp in length
47691: gap of unknown length
49059: contig of 1368 bp in length
49159: gap of unknown length
50579: contig of 1420 bp in length
50679: gap of unknown length
52301: contig of 1622 bp in length
52401: gap of unknown length

52402 53878: contig of 1477 bp in length
* 53879 53078: gap of unknown length
* 53979 55335: contig of 1257 bp in length
* 55336 55335: gap of unknown length
* 55336 56761: contig of 1426 bp in length
* 56762 56861: gap of unknown length
* 56762 58579: contig of 1718 bp in length
* 58580 60559: gap of unknown length
* 60560 60759: gap of unknown length
* 60760 62393: contig of 1534 bp in length
* 62394 62393: gap of unknown length
* 62394 63770: contig of 1377 bp in length
* 63771 63870: gap of unknown length
* 63871 65988: contig of 2028 bp in length
* 65989 65988: gap of unknown length
* 65999 68024: contig of 2026 bp in length
* 68025 68125: gap of unknown length
* 68125 69836: contig of 1612 bp in length
* 69837 69836: gap of unknown length
* 69837 71799: contig of 1963 bp in length
* 71800 71899: gap of unknown length
* 71900 73937: contig of 2038 bp in length
* 73938 74037: gap of unknown length
* 74038 76499: contig of 2462 bp in length
* 76500 76599: gap of unknown length
* 76500 78926: contig of 2227 bp in length
* 78927 81232: contig of 2306 bp in length
* 81233 81332: gap of unknown length
* 81333 83435: contig of 2103 bp in length
* 83436 83535: gap of unknown length
* 83536 86417: contig of 2882 bp in length
* 86418 86517: gap of unknown length
* 86518 88799: contig of 2282 bp in length
* 88800 91723: contig of 2824 bp in length
* 91724 91823: gap of unknown length
* 91824 94527: contig of 2704 bp in length
* 94528 94628: gap of unknown length
* 94628 98674: contig of 4047 bp in length
* 98675 98774: gap of unknown length
* 98775 100736: contig of 1962 bp in length
* 100737 100836: gap of unknown length
* 100837 104297: contig of 3451 bp in length

Query Match 0.6%; Score 21; DB 2: Length 168763;
Best Local Similarity 100.0%; Pred. NO. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2833 AAAACTTGCCTGAGCTCTTTC 2853
Db 13046 AAAACTTGCCTGAGCTCTTTC 13026

RESULT 37
AP001369/c
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-68317 map 18q12, WORKING
DRAFT SEQUENCE, 17 unordered pieces.
ACCESSION AP001369
VERSION AP001369.2 GI:8117282
KEYWORDS HTG; HTGS_PHASE1; HTG; HTGS_DRAFT
SOURCE Homo sapiens DNA, clone:RP11-68317.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171196)
Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,
Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.
TITLE Homo sapiens 171,196 genomic DNA of 18q12
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 171196)

AUTHORS Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,
Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT On May 30, 2000 this sequence version replaced gi:7229104.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-68317
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990129
Consensus quality: 159730 bases at least Q40
Consensus quality: 165678 bases at least Q30
Consensus quality: 168302 bases at least Q20
Insert size: 169596; sum-of-contigs
Quality coverage: 5.19x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
17 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 42291 contig of 42291 bp in length
42392 63147 contig of 20756 bp in length
63248 77851 contig of 14604 bp in length
77952 88544 contig of 10593 bp in length
88645 98724 contig of 10080 bp in length
98825 109077 contig of 10253 bp in length
109178 118993 contig of 9816 bp in length
119094 128219 contig of 9126 bp in length
128320 136675 contig of 8356 bp in length
136776 144488 contig of 7713 bp in length
144589 150077 contig of 5489 bp in length
150178 156798 contig of 6621 bp in length
156899 161857 contig of 4959 bp in length
161958 165379 contig of 3422 bp in length
165480 168313 contig of 2834 bp in length
168414 169434 contig of 1021 bp in length
169535 171196 contig of 1662 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 42291: contig of 42291 bp in length
42292 42391: gap of 100 bp
42392 63147: contig of 20756 bp in length
63148 63247: gap of 100 bp
63248 77851: contig of 14604 bp in length
77852 77951: gap of 100 bp
77952 88544: contig of 10593 bp in length
88545 88644: gap of 100 bp
88645 98724: contig of 10080 bp in length
98725 98824: gap of 100 bp
98825 109077: contig of 10253 bp in length
109078 109177: gap of 100 bp

AUTHORS
TITLE
JOURNAL
COMMENT

* 109178 118993: contig of 9816 bp in length
* 118994 119093: gap of 100 bp
* 119094 128219: contig of 9126 bp in length
* 128220 128319: gap of 100 bp
* 128320 136675: contig of 8356 bp in length
* 136676 136775: gap of 100 bp
* 136776 144488: contig of 7713 bp in length
* 144489 144588: gap of 100 bp
* 144589 150077: contig of 5489 bp in length
* 150078 150177: gap of 100 bp
* 150178 156798: contig of 6621 bp in length
* 156799 156898: gap of 100 bp
* 156899 161857: contig of 4959 bp in length
* 161858 161957: gap of 100 bp
* 161958 165379: contig of 3422 bp in length
* 165380 165479: gap of 100 bp
* 165480 168313: contig of 2834 bp in length
* 168314 168413: gap of 100 bp
* 168414 169434: contig of 1021 bp in length
* 169435 169534: gap of 100 bp
* 169535 171196: contig of 1662 bp in length.
FEATURES
 source
 1. .171196
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q12"
 /clone="RP11-68317"
 1. .42291
 /note="assembly_fragment"
 42392. .63147
 /note="assembly_fragment"
 63248. .77851
 /note="assembly_fragment"
 77952. .88544
 /note="assembly_fragment clone_end:SP6 vector_side:left"
 88645. .98724
 /note="assembly_fragment"
 98825. .109077
 /note="assembly_fragment"
 109178. .118993
 /note="assembly_fragment"
 119094. .128219
 /note="assembly_fragment"
 128320. .136675
 /note="assembly_fragment"
 136776. .144488
 /note="assembly_fragment"
 144589. .150077
 /note="assembly_fragment clone_end:T7 vector_side:right"
 150178. .156798
 /note="assembly_fragment"
 156899. .161857
 /note="assembly_fragment"
 161958. .165379
 /note="assembly_fragment"
 165480. .168313
 /note="assembly_fragment"
 168414. .169434
 /note="assembly_fragment"
 169535. .171196
 /note="assembly_fragment"
BASE COUNT 48636 a 35335 c 36221 g 49404 t 1600 others
ORIGIN

Query Match 0.6%; Score 21; DB 2; Length 171196;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 579 AATTTTTCCTTTTATCTA 599
|||||
Db 146855 AATTTTTCCTTTTATCTA 146835

RESULT 38
AP000880
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
 source
 1. .171853
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-839D17"
BASE COUNT 47378 a 38606 c 37154 g 48715 t
ORIGIN

Query Match 0.6%; Score 21; DB 9; Length 171853;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 579 AATTTTTCCTTTTATCTA 599
|||||
Db 51133 AATTTTTCCTTTTATCTA 51153

RESULT 39
AC126420/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Barna.N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,P., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna.N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,P., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,

AC126420
Mus musculus clone RP23-153M4, WORKING DRAFT SEQUENCE, 30 ordered
pieces.
AC126420
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 175245)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-153M4
Unpublished
2 (bases 1 to 175245)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna.N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,P., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,

McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Tesfaye, S., Theodore, J., Topman, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

COMMENT

Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Shit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

Project Information
 Center clone name: 153_M.4

Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 16436 bases at least Q40
 Consensus quality: 169284 bases at least Q30
 Consensus quality: 171255 bases at least Q20
 Insert size: 183000; agarose-fp
 Insert coverage: 172345; sum-of-contigs
 Quality coverage: 4.8 in Q20 bases; agarose-fp
 Quality coverage: 5.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 30 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces
 is believed to be correct as given, however the sizes
 of the gaps between them are based on estimates that have
 been provided by the submitter.
 This sequence will be replaced
 by the finished sequence as soon as it is available and
 the accession number will be preserved.

1 2956: contig of 2956 bp in length
 3057 3056: gap of 100 bp
 3057 4129: contig of 1073 bp in length
 4130 4229: gap of 100 bp
 4230 5714: contig of 1485 bp in length
 5715 5814: gap of 100 bp
 5815 7318: contig of 1504 bp in length
 7319 7418: gap of 100 bp
 7419 8623: contig of 1205 bp in length
 8624 8723: gap of 100 bp
 8724 10175: contig of 1452 bp in length
 10176 10275: gap of 100 bp
 10276 11370: contig of 1095 bp in length
 11371 11470: gap of 100 bp
 11471 12614: contig of 1144 bp in length
 12615 12714: gap of 100 bp
 12715 14823: contig of 2109 bp in length
 14824 14923: gap of 100 bp
 14924 16075: contig of 1152 bp in length
 16076 16175: gap of 100 bp
 16176 48108: contig of 31933 bp in length
 48109 48208: gap of 100 bp
 48209 50534: contig of 2326 bp in length
 50535 50634: gap of 100 bp
 50635 52640: contig of 2006 bp in length
 52641 52740: gap of 100 bp
 52741 55523: contig of 2783 bp in length
 55524 55623: gap of 100 bp
 55624 58689: contig of 3066 bp in length

58690 58789: gap of 100 bp
 58790 61160: contig of 2371 bp in length
 61161 61260: gap of 100 bp
 61261 6311: contig of 2051 bp in length
 6312 6341: gap of 100 bp
 6342 65922: contig of 2511 bp in length
 65923 66022: gap of 100 bp
 66023 68151: contig of 2129 bp in length
 68152 68251: gap of 100 bp
 68252 70510: contig of 2259 bp in length
 70511 70610: gap of 100 bp
 70611 72725: contig of 2115 bp in length
 72726 72825: gap of 100 bp
 72826 74380: contig of 1555 bp in length
 74381 74480: gap of 100 bp
 74481 77461: contig of 2981 bp in length
 77462 77561: gap of 100 bp
 77562 89014: contig of 11453 bp in length
 89015 89114: gap of 100 bp
 89115 100886: contig of 11772 bp in length
 100887 100986: gap of 100 bp
 100987 116990: contig of 16004 bp in length
 116991 117090: gap of 100 bp
 117091 135503: contig of 18413 bp in length
 135504 135603: gap of 100 bp
 135604 154923: contig of 19320 bp in length
 154924 155023: gap of 100 bp
 155024 174663: contig of 19640 bp in length
 174664 174763: gap of 100 bp
 174764 175245: contig of 482 bp in length.

FEATURES

Source

1. .175245
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-153M4"
 /clone_lib="RPC1-23 Female Mouse BAC"

misc_feature

1. .2956
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left

misc_feature

3057..4129
 /note="assembly_fragment"

misc_feature

4230..5714
 /note="assembly_fragment"

misc_feature

5815..7318
 /note="assembly_fragment"

misc_feature

7419..8623
 /note="assembly_fragment"

misc_feature

8724..10175
 /note="assembly_fragment"

misc_feature

10276..11370
 /note="assembly_fragment"

misc_feature

11471..12614
 /note="assembly_fragment"

misc_feature

12715..14823
 /note="assembly_fragment"

misc_feature

14924..16075
 /note="assembly_fragment"

misc_feature

16176..48108
 /note="assembly_fragment"

misc_feature

48209..50534
 /note="assembly_fragment"

misc_feature

50635..52640
 /note="assembly_fragment"

misc_feature

52741..55523
 /note="assembly_fragment"

misc_feature

55624..58689
 /note="assembly_fragment"

misc_feature

58790..61160
 /note="assembly_fragment"

misc_feature

61261..63111
 /note="assembly_fragment"

misc_feature

63412..65922

```

misc_feature      /note="assembly_fragment"
66023..68151
/note="assembly_fragment"
68252..70510
/note="assembly_fragment"
70611..72725
/note="assembly_fragment"
72826..74380
/note="assembly_fragment"
74481..77451
/note="assembly_fragment"
77562..89014
/note="assembly_fragment"
89115..100886
/note="assembly_fragment"
100987..116990
/note="assembly_fragment"
117091..135503
/note="assembly_fragment"
135604..154923
/note="assembly_fragment"
155024..174663
/note="assembly_fragment"
174764..175245
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT      51998 a 36415 c 35671 g 48251 t 2910 others
ORIGIN
Query Match      0.6%; Score 21; DB 2; Length 175245;
Best Local Similarity 100.0%; Pred. NO. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2915 AAAGAATACCCAGACTTCATG 2935
|||||
Db 81991 AAAGAATACCCAGACTTCATG 81971

RESULT 40
AP002840
LOCUS
DEFINITION      Homo sapiens genomic DNA, chromosome 11q, clone:RP11-159N11,
complete sequences.
ACCESSION      AP002840
VERSION        AP002840.2 GI:13488763
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Published Only in DataBase
2 (bases 1 to 177034)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (04-AUG-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Mar 29, 2001 this sequence version replaced gi:9757543.
COMMENT
FEATURES
Location/Qualifiers
1..177034
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-159N11"

```

```

BASE COUNT      48001 a 42063 c 41121 g 45849 t
ORIGIN
Query Match      0.6%; Score 21; DB 9; Length 177034;
Best Local Similarity 100.0%; Pred. NO. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1966 CTTGGATGGAGCAATATCA 1986
|||||
Db 81883 CTTGGATGGAGCAATATCA 81903

RESULT 41
AC009420
LOCUS
DEFINITION      Homo sapiens clone RP11-44J17, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION      AC009420
VERSION        AC009420.3 GI:14150911
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178229)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-44J17
Unpublished
2 (bases 1 to 178229)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funker,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tefaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 20, 2001 this sequence version replaced gi:7407920.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: Li079
Center clone name: 44_J-17
----- Summary Statistics
Sequencing vector: M13; M7815; 31% of reads
Sequencing vector: Plasmid; n/a; 69% of reads
Chemistry: Dye-primer-amersham; 2% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 175652 bases at least Q40
Consensus quality: 176863 bases at least Q30
Consensus quality: 177333 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 177629; sum-of-contigs
Quality coverage: 11.3 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces

```

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 51678: contig of 51678 bp in length
 * 51679 51778: gap of 100 bp
 * 51779 5422: contig of 644 bp in length
 * 52423 52522: gap of 100 bp
 * 52523 53221: contig of 699 bp in length
 * 53222 53321: gap of 100 bp
 * 53322 54413: contig of 1092 bp in length
 * 54414 54513: gap of 100 bp
 * 54514 55919: contig of 1406 bp in length
 * 55920 56019: gap of 100 bp
 * 56020 117029: contig of 61010 bp in length
 * 117030 117129: gap of 100 bp
 * 117130 178229: contig of 61100 bp in length.

FEATURES

source
 1. 178229
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-44J17"
 /clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. 51678
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"

misc_feature

51779..52422
 /note="assembly_fragment"

misc_feature

52523..53221
 /note="assembly_fragment"

misc_feature

53322..54413
 /note="assembly_fragment"

misc_feature

54514..55919
 /note="assembly_fragment"

misc_feature

56020..117029
 /note="assembly_fragment"

misc_feature

117130..178229
 /note="assembly_fragment"

clone_end:T7
 vector_side:right"

BASE COUNT 57859 33370 32611 g 53789 t 600 others
 ORIGIN

Query Match 0.6% Score 21; DB 2; Length 178229;

Best Local Similarity 100.08; Pred. No. 27;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3395 TTGAAAAGAGCTCATTCATT 3415

|||||
 Db 143209 TTGAAAAGAGCTCATTCATT 143229

RESULT 42

AC109179/c 179025 bp DNA linear HTG 28-APR-2002
 LOCUS Mus musculus clone RP23-382N7, WORKING DRAFT SEQUENCE, 14 ordered
 DEFINITION pieces.

AC109179 2 GT:20336065

VERSION HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS Mus musculus

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 179025)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

JOURNAL

REFERENCE 2 (bases 1 to 179025)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,

Seaman,S., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Chang,J., Chazaro,B.,
 Choepel,I., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fardo,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Hartas,A., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
 Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Stojanovic,N.,
 Severi,P., Spencer,B., Stange-Thomann,N., Tesfaye,S., Theodore,J.,
 Strauss,N., Subramanian,A., Talamas,J., Trigglio,J., Vassiliev,H.,
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

REFERENCE

AUTHORS

COMMENT

DIRECT SUBMISSION

SUBMITTED

RESEARCH

3 (bases 1 to 179025)

BIRREN,B., LINTON,L., NUSBAUM,C., LANDER,E., ALL,A., ALLEN,N.,

ANDERSON,S., BARN,N., BASTIEN,V., BLOOM,T., BOGUSLAVSKIY,L.,

BOUKHGALETER,B., BROWN,A., CAMARATA,J., CAMPOLANO,A., CHANG,J.,

CHAZARO,B., CHOPEL,I., COLANGELO,M., COLLINS,S., COLLYMORE,A.,

COOK,A., COOKE,P., DEARELLANO,K., DEWAR,K., DIAZ,J.S., DODGE,S.,

FARDO,S., FERREIRA,P., FITZHUGH,W., GAGE,D., GALAGAN,J., GARDYNA,S.,

GAGOS,B., GORD,S., GOYETTE,M., GRAHAM,L., GRAND-PIERRE,N.,

HAGOS,B., HORTON,L., HULME,W., ILIEV,I., JOHNSON,R., JONES,C.,

KAMAT,A., KARATAS,A., KELLIS,C., LAROCQUE,K., LAMAZARES,R.,

LANDERS,T., LEHOCZY,J., LEVINE,R., LINDBLAD-TOH,K., LIU,G.,

MACLEAN,C., MACDONALD,P., MAJOR,J., MARQUIS,N., MATTHEWS,C.,

MCARTHUR,M., MCEWAN,P., MCKERNAN,K., MELDRUM,J., MENEUS,L.,

MINOVA,I., MLENGA,V., MURPHY,T., NAYLOR,J., NGUYEN,C., NICOL,R.,

NORBU,C., NORMAN,C.H., O'CONNOR,T., O'DONNELL,P., O'NEIL,D.,

OLIVER,J., PETERSON,K., PHUNKHANG,P., PIERRE,N., POLLARA,V.,

RAYMOND,C., RETTA,R., RIEBACK,M., RILEY,R., RISE,C., ROGOV,P.,

ROMAN,J., ROSETTI,M., ROY,A., SANTOS,R., SCHAUER,S., SCHUPBACK,R.,

SEAMAN,S., SEVERI,P., SPENCER,B., STANGE-THOMANN,N., STOJANOVIC,N.,

STRAUSS,N., SUBRAMANIAN,A., TALAMAS,J., TESFAYE,S., THEODORE,J.,

TOPHAM,K., TRAVERS,M., TRAVIS,N., TRIGGLIO,J., VASSILIEV,H.,

VIEL,R., VO,A., WILSON,B., WU,X., WYMAN,D., YE,W.J., YOUNG,G.,

ZAINOUN,J., ZEMBOK,L., ZIMMER,A. AND ZODY,M.

DIRECT SUBMISSION

SUBMITTED (28-APR-2002)

RESEARCH, 320 CHARLES STREET, CAMBRIDGE, MA 02141, USA

ON APR 28, 2002 THIS SEQUENCE VERSION REPLACED GI:18482098.

ALL REPEATS WERE IDENTIFIED USING REPEATMASKER:

SMIT, A.F.A. & GREEN, P. (1996-1997)

HTTP://FTP.GENOME.WASHINGTON.EDU/RM/REPEATMASKER.HTML

----- GENOME CENTER

CENTER: WHITEHEAD INSTITUTE/ MIT CENTER FOR GENOME RESEARCH

CENTER CODE: WTBR

WEB SITE: HTTP://WWW.SEQ.WI.MIT.EDU

CONTACT: SEQUENCE.SUBMISSIONS@GENOME.WI.MIT.EDU

----- PROJECT INFORMATION

CENTER PROJECT NAME: L19614

CENTER CLONE NAME: 382_N-7

----- SUMMARY STATISTICS

SEQUENCING VECTOR: PLASMID; N/A; 100% OF READS

CHEMISTRY: DYE-TERMINATOR BIG DYE; 100% OF READS

ASSEMBLY PROGRAM: PHRAP; VERSION 0.960731

CONSENSUS QUALITY: 175637 BASES AT LEAST Q40

CONSENSUS QUALITY: 177170 BASES AT LEAST Q30

CONSENSUS QUALITY: 177548 BASES AT LEAST Q20

INSERT SIZE: 182000; AGAROSE-FP

INSERT SIZE: 177725; SUM-OF-CONTIGS

QUALITY COVERAGE: 7.4 IN Q20 BASES; AGAROSE-FP

QUALITY COVERAGE: 7.6 IN Q20 BASES; SUM-OF-CONTIGS

98425 104662 contig of 6238 bp in length
104763 110408 contig of 5646 bp in length
110509 115293 contig of 47815 bp in length
115394 120692 contig of 5299 bp in length
120793 125662 contig of 4770 bp in length
125663 130630 contig of 4968 bp in length
130731 133959 contig of 3229 bp in length
134060 136981 contig of 3501 bp in length
137082 140582 contig of 3389 bp in length
140683 144071 contig of 2847 bp in length
144172 147018 contig of 1377 bp in length
147119 148495 contig of 1925 bp in length
148596 151588 contig of 100 bp in length
151689 153261 contig of 100 bp in length
153362 154918 contig of 100 bp in length
154919 155018 contig of 100 bp in length
155019 156556 contig of 1538 bp in length
156557 156656 contig of 100 bp in length
156657 158209 contig of 1553 bp in length
158210 158309 contig of 100 bp in length
158310 160558 contig of 2249 bp in length
160559 160658 contig of 100 bp in length
160659 162643 contig of 1985 bp in length
162644 162743 contig of 100 bp in length
162744 164322 contig of 1579 bp in length
164323 164422 contig of 100 bp in length
164423 165625 contig of 1203 bp in length
165626 165725 contig of 100 bp in length
165726 166747 contig of 1022 bp in length
166748 166847 contig of 100 bp in length
166848 168335 contig of 1388 bp in length
168336 168335 contig of 100 bp in length
168336 169662 contig of 1327 bp in length
169663 169762 contig of 100 bp in length
169763 170933 contig of 1171 bp in length
170934 171033 contig of 100 bp in length
171034 172675 contig of 1642 bp in length
172676 172775 contig of 100 bp in length
172776 173786 contig of 1011 bp in length
173787 173886 contig of 100 bp in length
173887 175318 contig of 1432 bp in length
175319 175418 contig of 100 bp in length
175419 177028 contig of 1610 bp in length
177029 177128 contig of 100 bp in length
177129 178141 contig of 1013 bp in length
178142 178241 contig of 100 bp in length
178242 179281 contig of 1040 bp in length.

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

1 31182: contig of 31182 bp in length
31183 31282: gap of 100 bp
31283 47814: contig of 16532 bp in length
47815 47914: gap of 100 bp
47915 63599: contig of 15685 bp in length
63600 63699: gap of 100 bp
63700 78592: contig of 14893 bp in length
78593 78692: gap of 100 bp
78693 92264: contig of 13572 bp in length
92265 92364: gap of 100 bp
92365 98324: contig of 5960 bp in length
98325 98424: gap of 100 bp
98425 104662: contig of 6238 bp in length
104663 104762: gap of 100 bp
104763 110408: contig of 5646 bp in length
110409 110508: gap of 100 bp
110509 115293: contig of 4785 bp in length
115294 115393: gap of 100 bp
115394 120692: contig of 5299 bp in length
120693 120792: gap of 100 bp
120793 125562: contig of 4770 bp in length
125563 125662: gap of 100 bp
125663 130630: contig of 4968 bp in length
130631 130730: gap of 100 bp
130731 133959: contig of 3229 bp in length
133960 134059: gap of 100 bp
134060 136981: contig of 2922 bp in length
136982 137081: gap of 100 bp
137082 140582: contig of 3501 bp in length
140583 140682: gap of 100 bp
140683 144071: contig of 3389 bp in length
144072 144171: gap of 100 bp

144172 147018: contig of 2847 bp in length
147019 147118: gap of 100 bp in length
147119 148495: contig of 1377 bp in length
148496 148595: gap of 100 bp in length
148596 149563: contig of 968 bp in length
149564 149663: gap of 100 bp in length
149664 151588: contig of 1925 bp in length
151589 151688: gap of 100 bp in length
151689 153261: contig of 1573 bp in length
153262 153361: gap of 100 bp in length
153362 154918: contig of 1557 bp in length
154919 155018: gap of 100 bp in length
155019 156556: contig of 1538 bp in length
156557 156656: gap of 100 bp in length
156657 158209: contig of 1553 bp in length
158210 158309: gap of 100 bp in length
158310 160558: contig of 2249 bp in length
160559 160658: gap of 100 bp in length
160659 162643: contig of 1985 bp in length
162644 162743: gap of 100 bp in length
162744 164322: contig of 1579 bp in length
164323 164422: gap of 100 bp in length
164423 165625: contig of 1203 bp in length
165626 165725: gap of 100 bp in length
165726 166747: contig of 1022 bp in length
166748 166847: gap of 100 bp in length
166848 168335: contig of 1388 bp in length
168336 168335: gap of 100 bp in length
168336 169662: contig of 1327 bp in length
169663 169762: gap of 100 bp in length
169763 170933: contig of 1171 bp in length
170934 171033: gap of 100 bp in length
171034 172675: contig of 1642 bp in length
172676 172775: gap of 100 bp in length
172776 173786: contig of 1011 bp in length
173787 173886: gap of 100 bp in length
173887 175318: contig of 1432 bp in length
175319 175418: gap of 100 bp in length
175419 177028: contig of 1610 bp in length
177029 177128: gap of 100 bp in length
177129 178141: contig of 1013 bp in length
178142 178241: gap of 100 bp in length.
178242 179281: contig of 1040 bp in length.

FEATURES
Location/Qualifiers
1..179281
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-711G5"
1..31182
/note="assembly_fragment"
31283..47814
/note="assembly_fragment"
47915..63599
/note="assembly_fragment"
63700..78592
/note="assembly_fragment"
78693..92264
/note="assembly_fragment"
92365..98324
/note="assembly_fragment"
98425..104662
/note="assembly_fragment"
104763..110408
/note="assembly_fragment"
110509..115293
/note="assembly_fragment"
115394..120692
/note="assembly_fragment"

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

Query Match 0.6%; Score 21; DB 2; Length 179281;
Best Local Similarity 100.0%; Pred. No. 27;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2329 GGAATATGTCAGGTTTGT 2349
|||||
Db 94307 GGAATATGTCAGGTTTGT 94327

RESULT 44
AC083994
LOCUS
DEFINITION Homo sapiens clone RP11-222H22, WORKING DRAFT SEQUENCE, 13
unordered pieces.
AC083994
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 181265)
Birren.B., Linton.L., Nusbaum.C. and Lander.E.
Homo sapiens, clone RP11-222H22
Unpublished
2 (bases 1 to 181265)
Birren.B., Linton.L., Nusbaum.C., Lander.E., Abraham.H., Allen.N.,
Anderson.S., Barna.N., Bastien.V., Bedal.F., Boguslavsky.L.,
Boukhalter.B., Brown.A., Burkett.G., Campopiano.A., Castle.A.,
Choepe.Y., Colangelo.M., Collins.S., Collamore.A., Cooke.P.,
DeArellano.K., Dewar.K., Diaz.J.S., Dodge.S., Ferreira.P.,
FitHugh.W., Gage.D., Galagan.J., Gardyna.S., Ginde.S., Goyette.M.,
Graham.L., Grand-Pierre.N., Hagos.B., Headford.A., Horton.L.,
Iliev.I., Johnson.R., Jones.C., Kann.L., Karatas.A., Lacroque.K.,
Lamazares.R., Landers.T., Lehoczy.J., Levine.R., Liu.C., Liu.G.,
Macdonald.P., Marquis.N., McCarthy.M., McEwan.P., McKernan.K.,
McPheeters.R., Meldrim.J., Menues.L., Mihova.T., Mlenga.V.,
Morrow.J., Murphy.T., Naylor.J., Norman.C.H., O'Connor.T.,
O'Donnell.P., O'Neil.D., Olivari.T.M., Oliver.J., Peterson.K.,
Pierre.N., Pisani.C., Pollara.V., Raymond.C., Rieback.M., Riley.R.,
Rogov.P., Rothman.D., Roy.A., Santos.R., Schauer.S., Severy.P.,
Sougnuez.C., Spencer.B., Stange-Thomann.N., Stojanovic.N.,
Strauss.N., Subramanian.A., Talamas.J., Tesfaye.S., Theodore.J.,
Tirrell.A., Travers.M., Triglio.J., Vassiliev.H., Viel.R., Vo.A.,
Wilson.B., Wu.X., Wyman.D., Ye.W.J., Young.G., Zainoun.J.,
Zimmer.A. and Zody.M.
Direct Submission
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11472
Center clone name: 222_H-22
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 175722 bases at least Q40
Consensus quality: 178732 bases at least Q30
Consensus quality: 179651 bases at least Q20
Insert size: 189000; agarose-ff
Insert size: 180065; sum-of-contigs
Quality coverage: 5.4 in Q20 bases; agarose-ff
Quality coverage: 5.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2629: contig of 2629 bp in length
2630 2729: gap of 100 bp
2730 2903: contig of 174 bp in length
2904 3003: gap of 100 bp
3004 4169: contig of 1166 bp in length
4170 4269: gap of 100 bp
4270 9015: contig of 4746 bp in length
9016 9115: gap of 100 bp
9116 16346: contig of 7231 bp in length
16347 16446: gap of 100 bp
16447 24593: contig of 8153 bp in length
24600 24699: gap of 100 bp
24700 34073: contig of 9374 bp in length
34074 34173: gap of 100 bp
34174 48313: contig of 14140 bp in length
48314 48413: gap of 100 bp
48414 74955: contig of 26542 bp in length
74956 75055: gap of 100 bp
75056 103827: contig of 28772 bp in length
103828 103927: gap of 100 bp
103928 133045: contig of 29118 bp in length
133046 133145: gap of 100 bp
133146 157551: contig of 24406 bp in length
157552 157651: gap of 100 bp
157652 181265: contig of 23614 bp in length.

FEATURES
Location/Qualifiers
1..181265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-222H22"
/clone_lib="RPC1-11 Human Male BAC"
1..2629
/note="assembly_fragment"
clone_end:SP6
vector_side:left
2730..2903
/note="assembly_fragment"
3004..4169
/note="assembly_fragment"
4270..9015
/note="assembly_fragment"
9116..16346
/note="assembly_fragment"
16447..24599
/note="assembly_fragment"
24700..34073
/note="assembly_fragment"
34174..48313
/note="assembly_fragment"
48414..74955
/note="assembly_fragment"
75056..103827
/note="assembly_fragment"
103928..133045
/note="assembly_fragment"
133146..157551
/note="assembly_fragment"
157652..181265
/note="assembly_fragment"
clone_end:T7
vector_side:right
56264 a 34004 c 34091 g 55706 t 1200 others

BASE COUNT 56264 a 34004 c 34091 g 55706 t 1200 others
ORIGIN
Query Match 0.6%; Score 21; DB 2; Length 181265;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1941 AATCAGACACATAAGTTAG 1961
|||||
Db 117759 AATCAGACACATAAGTTAG 117779

RESULT 45

AC113269 183951 bp DNA linear HTG 06-JUN-2002
MUS musculus clone RP23-372F2, WORKING DRAFT SEQUENCE, 13 ordered
pieces.
AC113269
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
MUS musculus
house mouse.

REFERENCE

AC113269 3 GI:21327574
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
MUS musculus
house mouse.

REFERENCE

AC113269 3 GI:21327574
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
MUS musculus
house mouse.

REFERENCE

AC113269 3 GI:21327574
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
MUS musculus
house mouse.

REFERENCE

AC113269 3 GI:21327574
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
MUS musculus
house mouse.

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:21313780.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20332
Center clone name: 372_F.2
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 180116 bases at least Q40
Consensus quality: 181644 bases at least Q30
Consensus quality: 182175 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 182751; sum-of-contigs
Quality coverage: 7.9 in Q20 bases; agarose-fp
Quality coverage: 7.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 8760: contig of 8760 bp in length
* 8761 8860: gap of 100 bp
* 8861 10331: contig of 1471 bp in length
* 10332 10431: gap of 100 bp
* 10432 11525: contig of 1094 bp in length
* 11526 11625: gap of 100 bp
* 11626 13553: contig of 1928 bp in length
* 13554 13653: gap of 100 bp
* 13654 20393: contig of 6740 bp in length
* 20394 20493: gap of 100 bp
* 20494 25204: contig of 4711 bp in length
* 25205 25304: gap of 100 bp
* 25305 35649: contig of 10345 bp in length
* 35650 35749: gap of 100 bp
* 35750 55284: contig of 19535 bp in length
* 55285 55384: gap of 100 bp
* 55385 74290: contig of 18906 bp in length
* 74291 74390: gap of 100 bp
* 74391 98334: contig of 23944 bp in length
* 98335 98434: gap of 100 bp
* 98435 127342: contig of 28908 bp in length
* 127343 127443: gap of 100 bp
* 127443 148645: contig of 21203 bp in length
* 148646 148745: gap of 100 bp
* 148746 183951: contig of 35206 bp in length.
Location/Qualifiers
1. 183951
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-372F2"
/clone_lib="RPCI-23 Female Mouse BAC"
1. 8760
/note="assembly_fragment"
vector_side:left
clone_end:sp6
8861. 10331
/note="assembly_fragment"
10432. 11525
/note="assembly_fragment"

FEATURES

source
misc_feature
misc_feature
misc_feature


```
----- 2157 2187 4631 4551 3608 3588
----- 6879 6905 117 1567 1554
----- 15114 15028 3117 3741 3711
----- 2779 2817 2549 1952 1933
----- 65 65 182 3214 3186
----- 1945 1860 5443 17299 17528
----- 4326 4363 8852 3122 3186
----- 4652 4628 4125 362 3186
----- 187 187 933 4432 4389
----- 316 316 1529 2729 2795
----- 14460 14089 6315 1751 1744
----- 2663 2694 1564 1324 1304
----- 2093 2080 4331 3915 3919
----- 9930 9994 3370 3954 3919
----- 4370 4363 1951 1572 1554
----- 513 513 3853 8737 8662
----- 2015 1994 9178 8423 8376
----- 1536 1513 3482 1959 1933
----- 4928 4930 6091 36 36
----- 220 220 557 1853 1848
----- 683 683 3254 1219 1201
----- 2804 2817 2546 1474 1455
----- 881 889 6405 16039 15844
----- 9220 9368 45 2397 2405
----- 428 428 807 2145 2138
----- 7185 7290 894 13246 13238
----- 2562 2560 4325 1054 1089
----- 11284 11207 3049 438 3078
----- 9486 9368 780 399 399
----- 719 719 40 2662 2641
----- 1851 1860 2524 7352 7426
----- 3459 3459 3459 6142 6153
----- 354 354 926 926 931
----- 76 76 214 214 214
----- 1585 1585 2308 2308 2302
----- 1150 1150 7936 7936 7982
----- 2984 2984 695 695 695

Query Match 0.6%; Score 21; DB 9; Length 185148;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3395 TTGAAAGAGCTCATTTCATT 3415
|||||
Db 175598 TTGAAAGAGCTCATTTCATT 175618

RESULT 47
AC084367 186386 bp DNA linear HTG 30-OCT-2000
LOCUS Homo sapiens chromosome 2 clone RP11-359B5, WORKING DRAFT SEQUENCE,
DEFINITION 33 unordered pieces.
ACCESSION AC084367
VERSION AC084367.2 GI:11039565
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 186386)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 186386)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Oct 30, 2000 this sequence version replaced gi:11024936.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0359B05
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-primer ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 166829 bases at least Q40
Consensus quality: 173228 bases at least Q30
Consensus quality: 176024 bases at least Q20
Insert size: 204000; agarose-fp
Insert size: 182402; sum-of-contigs
Quality coverage: 3.47 in Q20 bases; agarose-fp
Quality coverage: 3.77 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
```


ORIGIN	REFERENCE
Query Match Best Local Similarity 0.6%; Score 21; DB 2; Length 186386; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	3 (bases 1 to 189748) Worley,K.C. Direct Submission Submitted (10-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA ----- Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GCCJ Center clone name: CH230-7B13 ----- Summary Statistics Sequencing vector: Plasmid; Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 125061 bases at least Q40 Consensus quality: 130947 bases at least Q30 Consensus quality: 136357 bases at least Q20 ----- * NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 66 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence. * as soon as it is available and the accession number will * be preserved.
QY 579 AATTTCCTTTCTTCTACTA 599	1 1015: contig of 1015 bp in length
	1115: gap of unknown length
DB 86483 AATTTCCTTTCTTCTACTA 86503	1016 2278: contig of 1163 bp in length
AC125777 189748 bp DNA linear HTG 10-JUL-2002	1116 2379: gap of unknown length
Rattus norvegicus clone CH230-7B13, *** SEQUENCING IN PROGRESS ***	1117 3604: contig of 1226 bp in length
66 unordered pieces.	1118 3704: gap of unknown length
AC125777	1119 4991: contig of 1287 bp in length
AC125777.1 GI:21630402	1120 5091: gap of unknown length
HTG: HTGS_PHASE1.	1121 5092: contig of 1138 bp in length
Norway rat.	1122 5329: gap of unknown length
Rattus norvegicus	1123 5717: contig of 1242 bp in length
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	1124 7672: gap of unknown length
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	1125 7672: gap of unknown length
1 (bases 1 to 189748)	1126 8696: contig of 1024 bp in length
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,	1127 8795: gap of unknown length
Albrooks,S.L., Amarantunge,H.C., Ate,J.R., Ayele,M., Banks,T.,	1128 8796: contig of 1030 bp in length
Barbaria,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,	1129 9826: gap of unknown length
Bouck,D., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P.,	1130 9826: contig of 1429 bp in length
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,	1131 11355: gap of unknown length
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,	1132 11455: contig of 1618 bp in length
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,	1133 13072: contig of 1337 bp in length
Cleveland,C.D., Cox,C., Coyle,M.C., Dathorne,S.R., David,R.,	1134 13173: contig of 1746 bp in length
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,	1135 14918: gap of unknown length
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinoh,K.H.,	1136 15018: gap of unknown length
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,	1137 16055: contig of 1037 bp in length
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,	1138 16056: gap of unknown length
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,	1139 16156: contig of 1444 bp in length
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,	1140 17600: gap of unknown length
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,	1141 17600: contig of 1337 bp in length
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,	1142 19037: gap of unknown length
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,	1143 19137: contig of 1283 bp in length
Homs,I., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,	1144 20420: gap of unknown length
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,	1145 20520: contig of 1071 bp in length
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,	1146 21591: gap of unknown length
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,	1147 21591: contig of 1372 bp in length
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounsegh,H.,	1148 21691: gap of unknown length
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,	1149 23163: contig of 1205 bp in length
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,	1150 23163: gap of unknown length
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,	1151 24368: contig of 1352 bp in length
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,	1152 25820: gap of unknown length
Moser,M., Neal,D., Newton,J., Newsham,N., Nguyen,A., Nguyen,N.,	1153 25920: contig of 1364 bp in length
Nguyen,N., Nickerson,E., Nwokwuo,S., Oguh,M., Okwuonu,G.,	1154 27384: gap of unknown length
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,	1155 27384: contig of 1245 bp in length
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,	
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,	
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,	
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,	
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,	
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,	
Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,	
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,	
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,	
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,	
Weinstock,G. and Gibbs,R.	
Unpublished	
Direct Submission	
2 (bases 1 to 189748)	
Worley,K.C.	
Direct Submission	
Submitted (30-JUN-2002) Human Genome Sequencing Center, Department	
of Molecular and Human Genetics, Baylor College of Medicine, One	
Baylor Plaza, Houston, TX 77030, USA	

Center: Baylor College of Medicine	
Center code: BCM	
Web site: http://www.hgsc.bcm.tmc.edu/	
Contact: hgsc-help@bcm.tmc.edu	

Project Information	
Center project name: GCCJ	
Center clone name: CH230-7B13	

Summary Statistics	
Sequencing vector: Plasmid;	
Chemistry: Dye-terminator Big Dye; 100% of reads	
Assembly program: Phrap; version 0.990329	
Consensus quality: 125061 bases at least Q40	
Consensus quality: 130947 bases at least Q30	
Consensus quality: 136357 bases at least Q20	

* NOTE: Estimated insert size may differ from sequence length	
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 66 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence. * as soon as it is available and the accession number will * be preserved.	


```

* 28529 28728: gap of unknown length
* 28729 30107: contig of 1379 bp in length
* 30208 30207: gap of unknown length
* 30208 31892: contig of 1685 bp in length
* 31893 31932: gap of unknown length
* 31933 33557: contig of 1565 bp in length
* 33558 33657: gap of unknown length
* 33658 34822: contig of 1165 bp in length
* 34823 34922: gap of unknown length
* 34923 36315: contig of 1393 bp in length
* 36316 36415: gap of unknown length
* 36416 38548: contig of 2133 bp in length
* 38549 38648: gap of unknown length
* 38649 40536: contig of 1888 bp in length
* 40537 40636: gap of unknown length
* 40637 41892: contig of 1256 bp in length
* 41893 41992: gap of unknown length
* 41993 43516: contig of 1524 bp in length
* 43517 43617: gap of unknown length
* 43617 44992: contig of 1376 bp in length
* 44993 45092: gap of unknown length
* 45093 46471: contig of 1379 bp in length
* 46472 46571: gap of unknown length
* 46572 48668: contig of 2097 bp in length
* 48669 48768: gap of unknown length
* 48769 50172: contig of 1404 bp in length
* 50173 50272: gap of unknown length
* 50273 53281: contig of 3009 bp in length
* 53282 53381: gap of unknown length
* 53382 55305: contig of 1924 bp in length
* 55306 55405: gap of unknown length
* 55406 56856: contig of 1451 bp in length
* 56857 56956: gap of unknown length
* 56957 59361: contig of 2405 bp in length
* 59362 59461: gap of unknown length
* 59462 61457: contig of 1996 bp in length
* 61458 61557: gap of unknown length
* 61558 63257: contig of 1700 bp in length
* 63258 63357: gap of unknown length
* 63358 65398: contig of 2041 bp in length
* 65399 65498: gap of unknown length
* 65499 69630: contig of 4132 bp in length
* 69631 69730: gap of unknown length
* 69731 72934: contig of 3204 bp in length
* 72935 73034: gap of unknown length
* 73035 76310: contig of 3276 bp in length
* 7311 76410: gap of unknown length
* 76411 78035: contig of 1625 bp in length
* 78036 78135: gap of unknown length
* 78136 80493: contig of 2358 bp in length
* 80494 80593: gap of unknown length
* 80594 83362: contig of 2769 bp in length
* 83363 83462: gap of unknown length
* 83463 85944: contig of 2382 bp in length
* 85945 89228: contig of 3284 bp in length
* 89229 89328: gap of unknown length
* 89329 92715: contig of 3387 bp in length
* 92716 92815: gap of unknown length
* 92816 96224: contig of 3409 bp in length
* 96225 96324: gap of unknown length
* 96325 96633: contig of 3309 bp in length
* 96634 99734: gap of unknown length
* 99734 103914: contig of 4181 bp in length
* 103915 104014: gap of unknown length
* 104015 109640: contig of 5626 bp in length
* 109641 109740: gap of unknown length

```

Query Match 0.6%; Score 21; DB 2; Length 189748;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2988 TCAGGAAGTGAAGGACAAAG 3008

||||| TCAGGAAGTGAAGGACAAAG 97956

RESULT 49 AC009466

LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS

AC009466 193279 bp DNA linear PRI 15-DEC-2000
 Homo sapiens chromosome 11, clone RP11-87N22, complete sequence.

HTG.
 AC009466.17 GI:11321807

SOURCE
 ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

1 (bases 1 to 193279)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 11, clone RP11-87N22
 Unpublished

2 (bases 1 to 193279)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
 Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
 Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
 Funk,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,B., Grant,G.,
 Karas,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kuan,L.,
 Karas,B., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
 Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
 Direct Submission

Submitted (24-AUG-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 193279)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Barna,N., Bastien,V., Bonuslavsky,L., Boukhalter,B., Brown,A.,
 Camarata,J., Campopiano,A., Chospel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
 Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
 Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
 Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
 McPheeters,R., Meldrum,J., Meneus,L., Mihova,I., Mienga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
 Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
 Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (15-DEC-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 23, 2000 this sequence version replaced gi:1181808.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information

TITLE JOURNAL REFERENCE AUTHORS

Submitted (15-DEC-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 23, 2000 this sequence version replaced gi:1181808.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information

Center project name: Li361
Center clone name: 87_N_22

FEATURES

Source

Location/Qualifiers

```

1..193279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-87N22"
/clone_lib="RPC1-11 Human Male BAC"
504..561
/rpt_family="(CA)n"
complement(573..882)
/rpt_family="AluY"
988..1015
/rpt_family="(TG)n"
complement(1077..1265)
/rpt_family="MIR"
complement(1319..1361)
/rpt_family="MIR"
complement(1646..1835)
/rpt_family="MIR"
complement(1943..2320)
/rpt_family="THELC"
complement(3451..3717)
/rpt_family="AluSg"
3852..3981
/rpt_family="MIR"
4678..4873
/rpt_family="L1PA3"
4907..4926
/rpt_family="(A)n"
5590..5758
/rpt_family="MIR3"
7409..7450
/rpt_family="T-rich"
9037..9083
/rpt_family="L2"
11416..11436
/rpt_family="AT-rich"
13119..13949
/rpt_family="L2"
complement(14968..15042)
/rpt_family="L2"
15058..15122
/rpt_family="L2"
complement(15282..15441)
/rpt_family="MIR"
15848..15899
/rpt_family="(CA)n"
complement(17151..17264)
/rpt_family="MER94"
17719..17739
/rpt_family="AT-rich"
17769..17839
/rpt_family="T-rich"
18599..18853
/rpt_family="L2"
complement(21244..21528)
/rpt_family="AluY"
complement(22197..22307)
/rpt_family="L3"
complement(22450..22875)
/rpt_family="L3"
complement(22884..23008)
/rpt_family="MER20B"
complement(23232..23349)
/rpt_family="L3"
complement(25680..25985)
/rpt_family="AluY"
29025..35170
/rpt_family="L1PA5"

```

```

repeat_region 36356..36375
/rpt_family="(CTG)n"
complement(39370..39845)
/rpt_family="L1MB3"
complement(39876..40070)
/rpt_family="L1M4"
42933..43060
/rpt_family="(TA)n"
43089..43144
/rpt_family="AT-rich"
complement(44577..44696)
/rpt_family="AluYa5/8"
46470..46529
/rpt_family="MIR"
47072..47286
/rpt_family="MIR"
47416..47698
/rpt_family="AluSx"
49722..49765
/rpt_family="(TG)n"
50251..50528
/rpt_family="AluJb"
complement(56872..56178)
/rpt_family="AluSx"
57210..57502
/rpt_family="AluJo"
57596..57678
/rpt_family="L2"
complement(57714..58028)
/rpt_family="AluSx"
58029..58051
/rpt_family="AT-rich"
complement(58542..58602)
/rpt_family="MIR3"
58932..59045
/rpt_family="MER81"
complement(59500..59545)
/rpt_family="MERSA"
60676..60695
/rpt_family="(A)n"
complement(61438..61631)
/rpt_family="MIR"
62350..62580
/rpt_family="MIR"
62788..62827
/rpt_family="AT-rich"
63566..63867
/rpt_family="L1MA4"
63934..64117
/rpt_family="MIR"
complement(64123..64428)
/rpt_family="AluYa5"
complement(64779..65078)
/rpt_family="AluSx"
67738..67813
/rpt_family="CT-rich"
complement(71638..71784)
/rpt_family="MERSA"
complement(71960..72305)
/rpt_family="L3"

```

Query Match 0.6%; Score 21; DB 9; Length 193279;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
|||||

DB 67736 AATTTTTCCTTTTATCTA 67756
|||||

RESULT 50
AC055817
LOCUS

AC055817

202781 bp DNA

linear HTG 06-AUG-2002

```

DEFINITION Mus musculus clone RP23-154I7, *** SEQUENCING IN PROGRESS ***, 8
unordered pieces.
ACCESSION AC055817
VERSION AC055817.5 GI:22123678
KEYWORDS HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birtten,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 202781)
Mus musculus, clone RP23-154I7
Unpublished
2 (bases 1 to 202781)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,K., Johnson,R., Jones,C., Kana,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 202781)
Birtten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camara,J., Chang,J., Chazaro,B., Choepel,Y., Collimore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:14140334.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8049
Center clone name: 154_1_7
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 32467: contig of 32467 bp in length
* 32468 32567: gap of 100 bp
* 32568 66596: contig of 34029 bp in length
* 66597 66696: gap of 100 bp
* 66697 88421: contig of 21725 bp in length
* 88422 88521: gap of 100 bp
* 88522 93034: contig of 4513 bp in length
* 93035 93134: gap of 100 bp
* 93135 131293: contig of 38159 bp in length
* 131294 131393: gap of 100 bp
* 131394 178881: contig of 47488 bp in length
* 178882 178981: gap of 100 bp
* 178982 201438: contig of 22457 bp in length
* 201439 201538: gap of 100 bp
* 201539 202781: contig of 1243 bp in length.

```

FEATURES

```

source
1..202781
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-154I7"
/clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT 56129 a 44477 c 44824 g 56482 t 869 others
ORIGIN

```

```

Query Match 0.6%; Score 21; DB 2; Length 202781;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 2694 TGGACCATGATGTCACAAATTG 2714

Db 168432 TGGACCATGATGTCACAAATTG 168452

Search completed: November 6, 2002, 05:43:43
Job time : 22032 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 19:30:52 ; Search time 705 Seconds

(without alignments)
11918.022 Million cell updates/sec

Title: US-09-782-874-1

Perfect score: 3731

Sequence: 1 GAATATCTTTACTTCTT.....AGTTTCATCTTCTCTCTAAA 3731

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : N_Geneseq_101002.*

```
1: /SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3731	100.0	3731	22	AAD04370 Tomato RNA-directe
2	3731	100.0	3731	23	AAS17837 Tomato RNA-directe
3	3495	93.7	3731	23	AAS17845 Tomato RNA-directe
C 4	30	0.8	906	17	AAT18078 Potato calmodulin
C 5	25	0.7	25	22	AAD04371 RACE PCR primer, G
C 6	25	0.7	25	22	AAD04373 RACE PCR primer, G
C 7	25	0.7	25	23	AAS17838 Tomato RNA-directe
C 8	25	0.7	25	23	AAS17840 Tomato RNA-directe
C 9	24	0.6	25	22	AAD04372 RACE PCR primer, G

C 10	24	0.6	25	23	AAS17839 Tomato RNA-directe
C 11	20	0.5	407	24	ABL64204 Stomach cancer rel
C 12	20	0.5	786	24	ABQ55295 Human ovarian anti
C 13	20	0.5	1000	24	AAD38196 Non-typeable Haemo
C 14	20	0.5	1509	20	AAZ52915 Human prostate tum
C 15	20	0.5	1730	20	AAZ18199 TIP49 coding seque
C 16	20	0.5	1744	21	AAZ59220 Human secreted pro
C 17	20	0.5	1750	20	AAZ24059 Human regulatory m
C 18	20	0.5	38653	22	AAZ72348 Human immune/haema
C 19	20	0.5	34	22	AAD04374 Forward PCR primer
C 20	19	0.5	34	23	AAS17841 Tomato RNA-directe
C 21	19	0.5	255	22	AAH81903 Rat differential t
C 22	19	0.5	327	24	ABN67264 Streptococcus poly
C 23	19	0.5	369	21	AAZ31197 Human secreted pro
C 24	19	0.5	452	22	AA187435 Human polynucleoti
C 25	19	0.5	456	18	AAZ49104 Partial DNA clone
C 26	19	0.5	461	22	ABA59152 Human foetal liver
C 27	19	0.5	461	22	AAK07338 Human brain expres
C 28	19	0.5	461	22	AAK33106 Human bone marrow
C 29	19	0.5	461	22	AAI38901 Human genome-deriv
C 30	19	0.5	461	24	AB507923 Human breast cance
C 31	19	0.5	757	22	AAZ22601 Arabidopsis thalia
C 32	19	0.5	783	21	AAC38539 Human neuroblastom
C 33	19	0.5	889	22	AA194184 Mouse pheromone re
C 34	19	0.5	1296	20	AAK05849 Human secreted pro
C 35	19	0.5	1405	21	AAA61289 Mouse pheromone re
C 36	19	0.5	1521	20	AAK05850 Drosophila melanog
C 37	19	0.5	2606	23	AB108786 Mouse pheromone re
C 38	19	0.5	2732	20	AAK05810 Mouse pheromone re
C 39	19	0.5	2962	20	AAK05811 DNA encoding a DNA
C 40	19	0.5	4446	19	AAV36034 Human musculoskele
C 41	19	0.5	4656	22	AA136325 Human small cell l
C 42	19	0.5	4877	22	AAD11120 Drosophila ORF1553
C 43	19	0.5	4924	21	ABL75998 Drosophila melanog
C 44	19	0.5	5415	23	ABL30262 Shrimp white spot
C 45	19	0.5	18234	22	AAH62719 Reverse PCR primer
C 46	19	0.5	305107	22	AAD04376 Tomato RNA-directe
C 47	19	0.5	35	23	AAS17843 Primer base sequen
C 48	18	0.5	35	23	AAS17843
C 49	18	0.5	48	22	AAF29312
C 50	18	0.5	48	22	AAF29312

ALIGNMENTS

RESULT 1

AAD04370

ID AAD04370 standard; cDNA; 3731 BP.

XX AAD04370;

AC AAD04370;

DT 04-JUL-2001 (first entry)

XX Tomato RNA-directed RNA polymerase (RdRP) cDNA.

XX Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein; ss.
XX Lycopersicon esculentum.

XX Location/Qualifiers

PH Key 194..3538

FT CDS /**tag=a

FT /product="Tomato C-protein having RNA-directed RNA
polymerase (RdRP) activity"

XX US6218142-B1.

XX 17-APR-2001.

XX 05-MAR-1997; 97US-0811583.

XX

PR 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.

PA (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-289830/30.

XX DR P-PSDB; AAE00897.

XX

XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA

XX polymerase enzymatic activity, useful in modulating gene expression in

XX plants, humans and animals, as well as in plant cell/tissue cultures or

XX plant breeding

XX

XX Claim 1; Column 25-34; 31pp; English.

XX

XX The present sequence is a cDNA encoding tomato C-protein having

XX of RNA-directed RNA polymerase (RdRP) activity. This protein is capable

XX synthesizing complementary RNA molecules. RdRP nucleic acid is useful

XX for modulating gene expression in plants, humans and animals. This may

XX lead to various physiological, developmental and/or morphological

XX changes. Transgenic plants containing RdRP nucleic acid is especially

XX useful in plant cell or tissue cultures and in plant breeding. RdRP is

XX useful in gene therapy, particularly for treating a disease that is

XX caused by the undesirable expression or overexpression of a gene.

XX

XX SQ Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 other;

XX

XX Query Match 100.0%; Score 3731; DB 22; Length 3731;

XX Best Local Similarity 100.0%; Pred. No. 0;

XX Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX 1 GAATATATCTTACTTACTTACACAGGATTTGACTTCACTCCCTCAAGTCCTTGTGT 60

XX

XX 1 GAATATATCTTACTTACTTACACAGGATTTGACTTCACTCCCTCAAGTCCTTGTGT 60

XX

XX 61 GTTGTGATAAATTTGGTGTGCTTCAGTTTCAGTCTACTGCTGGGTAGTTTAT 120

XX

XX 61 GTTGTGATAAATTTGGTGTGCTTCAGTTTCAGTCTACTGCTGGGTAGTTTAT 120

XX

XX 121 TTTCATTAATCTAGGGGATTTCCAGTTGGTGTAGCAATTTGAAGTCGAATCGACTT 180

XX

XX 121 TTTCATTAATCTAGGGGATTTCCAGTTGGTGTAGCAATTTGAAGTCGAATCGACTT 180

XX

XX 181 GGAATTTGGCTACATGGGAAGCAATTCAGGTTTTCGGATTCCCTTATCTCTCTGC 240

XX

XX 181 GGAATTTGGCTACATGGGAAGCAATTCAGGTTTTCGGATTCCCTTATCTCTCTGC 240

XX

XX 241 GGAAGTGTAAAGTCATCTAGAGAAATATACAGATATGGAATCTGATGCAATGGA 300

XX

XX 241 GGAAGTGTAAAGTCATCTAGAGAAATATACAGATATGGAATCTGATGCAATGGA 300

XX

XX 301 GGTAAACAGTCCAAAGGAGATCTAGAGCAATTTGCCAAGTCAATTTGCCGACAAT 360

XX

XX 301 GGTAAACAGTCCAAAGGAGATCTAGAGCAATTTGCCAAGTCAATTTGCCGACAAT 360

XX

XX 361 AAGTGTGACAAATATCATCTTGGCTAATACAGCTGTATTTGGCTCTCTTATTT 420

XX

XX 361 AAGTGTGACAAATATCATCTTGGCTAATACAGCTGTATTTGGCTCTCTTATTT 420

XX

XX 421 GAAGGCTTGGGAATGAATGATATTTGCCAATTCGCGGCATATGCGATGGA 480

XX

XX 421 GAAGGCTTGGGAATGAATGATATTTGCCAATTCGCGGCATATGCGATGGA 480

XX

XX 481 TGGCATTAATTTGGATTCGAGTGCAGATATCAGATGACAAGTTGCAAGTCTGGGAG 540

XX

XX 481 TGGCATTAATTTGGATTCGAGTGCAGATATCAGATGACAAGTTGCAAGTCTGGGAG 540

XX

XX 541 TACAGAAGTTTCAATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCAT 600

XX

XX 541 TACAGAAGTTTCAATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCAT 600

XX

XX

XX

XX

QY 601 TGGTTAGCTGACTATAAAGCTTCACTTTCATATGAAAATATATGCAGGTTGTCTCCA 660

Db 601 TGGTTAGCTGACTATAAAGCTTCACTTTCATATGAAAATATATGCAGGTTGTCTCCA 660

QY 661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTTATTTGGTGTCTCCGGAT 720

Db 661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTTATTTGGTGTCTCCGGAT 720

QY 721 CTATAGAGACTTTGAACCTCCTGTTATAGCTTCTTTAAGGAACCTCCTGATGATCAGTG 780

Db 721 CTATAGAGACTTTGAACCTCCTGTTATAGCTTCTTTAAGGAACCTCCTGATGATCAGTG 780

QY 781 GGTGAGACAAACAGATTTCCCTCCATCTTTGGATAGGGCTATCTTCTAGCTTATGTTTGA 840

Db 781 GGTGAGACAAACAGATTTCCCTCCATCTTTGGATAGGGCTATCTTCTAGCTTATGTTTGA 840

QY 841 GTTCCGTAGGGTGTGCTTCCAAATTTCCAGGAAGTTTTCCTCAGTATGCAGAACG 900

Db 841 GTTCCGTAGGGTGTGCTTCCAAATTTCCAGGAAGTTTTCCTCAGTATGCAGAACG 900

QY 901 TGAACAAATATTTACTTTACAGACTGGTTTCCACCTTTTCGCTCTCAAAAATCGGCTCT 960

Db 901 TGAACAAATATTTACTTTACAGACTGGTTTCCACCTTTTCGCTCTCAAAAATCGGCTCT 960

QY 961 GGTCCCAATGTCCAGCTCCGGAAGAAATTTCAATTCCTCCTACAGATTTTGTCTCAAAAT 1020

Db 961 GGTCCCAATGTCCAGCTCCGGAAGAAATTTCAATTCCTCCTACAGATTTTGTCTCAAAAT 1020

QY 1021 TAGTCTTTTGGTACAGCATGGATGTCATACCTGGGCGAGCAATTAATGCTACTTTTCGG 1080

Db 1021 TAGTCTTTTGGTACAGCATGGATGTCATACCTGGGCGAGCAATTAATGCTACTTTTCGG 1080

QY 1081 ATTAGTTGATCCTCGAAGGAGAAATGTCGATGTCATGAGCATGCTTAGAGAAACTGTA 1140

Db 1081 ATTAGTTGATCCTCGAAGGAGAAATGTCGATGTCATGAGCATGCTTAGAGAAACTGTA 1140

QY 1141 CTATATAAGAGATGCTGTTATGATCCCGTAGGTTGGCTCAGTACGATGATGATGATGATG 1200

Db 1141 CTATATAAGAGATGCTGTTATGATCCCGTAGGTTGGCTCAGTACGATGATGATGATGATG 1200

QY 1201 TCTCAAGGTPAGCAACCTCCAAAATCTCCGTCATCTTACATTTAGATGATGATGATGATG 1260

Db 1201 TCTCAAGGTPAGCAACCTCCAAAATCTCCGTCATCTTACATTTAGATGATGATGATGATG 1260

QY 1261 TGTAAAGAGGTCCTAGTACACCATGCAAGTTTATTTTGTGTCAGAGGTTTAAATGT 1320

Db 1261 TGTAAAGAGGTCCTAGTACACCATGCAAGTTTATTTTGTGTCAGAGGTTTAAATGT 1320

QY 1321 TTCCAATCGGGTCTCCGCAATTTATCTGAAGACATAGATAAATCTTCTCGTGTCTTT 1380

Db 1321 TTCCAATCGGGTCTCCGCAATTTATCTGAAGACATAGATAAATCTTCTCGTGTCTTT 1380

QY 1381 TGTGATGAGGATGGGAGAACTGTTTCTACAGACTTATACCAAGCAAGTACTGG 1440

Db 1381 TGTGATGAGGATGGGAGAACTGTTTCTACAGACTTATACCAAGCAAGTACTGG 1440

QY 1441 AAGTGTGTTCAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTGT 1500

Db 1441 AAGTGTGTTCAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTGT 1500

QY 1501 AATTGGTGATAAATAATTTGAATTTCTTGCATTTTCAATGAGCCAGTTGCGGGATTAATTC 1560

Db 1501 AATTGGTGATAAATAATTTGAATTTCTTGCATTTTCAATGAGCCAGTTGCGGGATTAATTC 1560

QY 1561 AGTGTGATTTTGGATCAAGACTGGCTTACTGCAATATGATATAAGAGCTTGGATGGG 1620

Db 1561 AGTGTGATTTTGGATCAAGACTGGCTTACTGCAATATGATATAAGAGCTTGGATGGG 1620

QY 1621 TGAATTTTCCAGATCAAGAAATGTCGAAAATATGCTGCCAGACTTGTCTAATCTTTGG 1680

Db 1621 TGAATTTTCCAGATCAAGAAATGTCGAAAATATGCTGCCAGACTTGTCTAATCTTTGG 1680

```
QY 1681 TTCTCCAGAGACGTTTGAGTGTCTTATAGGCATGAGATTGAAGTTATTCCTCCGATGATAA 1740
Db 1681 TTCTCCAGAGACGTTTGAGTGTCTTATAGGCATGAGATTGAAGTTATTCCTCCGATGATAA 1740
QY 1741 GGTTCATGGAACAGCTATGCTTTTCTGATGGAATTGGTAAATATCTGGTGACTTTC 1800
Db 1741 GGTTCATGGAACAGCTATGCTTTTCTGATGGAATTGGTAAATATCTGGTGACTTTC 1800
QY 1801 TCATAGAGTTGCCCTCAAAATGTGGCTTCAATATACCCCATCTGCTTCCAGATTGCTTA 1860
Db 1801 TCATAGAGTTGCCCTCAAAATGTGGCTTCAATATACCCCATCTGCTTCCAGATTGCTTA 1860
QY 1861 TGTGGATATAAGGTGTGTGGGTGTGATCCGGATTCATCAATCAAGTTGCTTTGAG 1920
Db 1861 TGTGGATATAAGGTGTGTGGGTGTGATCCGGATTCATCAATCAAGTTGCTTTGAG 1920
QY 1921 AAAGAGCATGCGAAATATGAATCAGACAACATAAAGTTAGATGCTCTTGGATGGAGCAA 1980
Db 1921 AAAGAGCATGCGAAATATGAATCAGACAACATAAAGTTAGATGCTCTTGGATGGAGCAA 1980
QY 1981 ATATCAGCCTTGTTATCTTAATGCTCAACCTGATTACGCTCTTCTACACTTGGAGTGA 2040
Db 1981 ATATCAGCCTTGTTATCTTAATGCTCAACCTGATTACGCTCTTCTACACTTGGAGTGA 2040
QY 2041 AGATGAAGTTCTCGAACAGAGCAAGAAAGAGCTGTAGATCAGCTTGATCTATCTTGCA 2100
Db 2041 AGATGAAGTTCTCGAACAGAGCAAGAAAGAGCTGTAGATCAGCTTGATCTATCTTGCA 2100
QY 2101 TGATCTTTGAAGCAGCAGAGGCTTTGGAATTGATGTCCTCGGAGAACACTAATAT 2160
Db 2101 TGATCTTTGAAGCAGCAGAGGCTTTGGAATTGATGTCCTCGGAGAACACTAATAT 2160
QY 2161 TCTCAAGCCTGCTAACTGCTGTTATAAGCCTGATGCTGAGCCCTTTCTTCAATGAT 2220
Db 2161 TCTCAAGCCTGCTAACTGCTGTTATAAGCCTGATGCTGAGCCCTTTCTTCAATGAT 2220
QY 2221 GTTGCAACCTTCGCGCATCCCAAGTTGCTGATTTGCGGACTAGATCAAGATATTTAT 2280
Db 2221 GTTGCAACCTTCGCGCATCCCAAGTTGCTGATTTGCGGACTAGATCAAGATATTTAT 2280
QY 2281 TCCAAATGGAAGAACATGATGGATGTTGGATGAATCCAGAACCTTGGAAATGGTCA 2340
Db 2281 TCCAAATGGAAGAACATGATGGATGTTGGATGAATCCAGAACCTTGGAAATGGTCA 2340
QY 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTTCCTGACGANTTACATCCATT 2400
Db 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTTCCTGACGANTTACATCCATT 2400
QY 2401 TAATAACAGAGATCCACCACAGTAATTTCAATCTCAAGGGAATCTGGTTGTCAAA 2460
Db 2401 TAATAACAGAGATCCACCACAGTAATTTCAATCTCAAGGGAATCTGGTTGTCAAA 2460
QY 2461 AAATCCATGTTGATCTGCTGATATTCGTTGTTTAAAGGCTGTAAATGTTGAGCGCT 2520
Db 2461 AAATCCATGTTGATCTGCTGATATTCGTTGTTTAAAGGCTGTAAATGTTGAGCGCT 2520
QY 2521 GCACCATAGGTAGATTGTTGATTTCCCTCAGAAAGGAAAGACCTCATCCGATGA 2580
Db 2521 GCACCATAGGTAGATTGTTGATTTCCCTCAGAAAGGAAAGACCTCATCCGATGA 2580
QY 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGATCAAGACATGAT 2640
Db 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGATCAAGACATGAT 2640
QY 2641 CCCGCCAAGCAAGTCCAGCGGATGGAATATCCTCCAGCACCAGCATACAGTTGGACCA 2700
Db 2641 CCCGCCAAGCAAGTCCAGCGGATGGAATATCCTCCAGCACCAGCATACAGTTGGACCA 2700
QY 2701 TGATGTCACAATTCAGGAAGTTGAAGTACTTCCACCACCTATATTTGTAATGACAGTTT 2760
Db 2701 TGATGTCACAATTCAGGAAGTTGAAGTACTTCCACCACCTATATTTGTAATGACAGTTT 2760
QY 2761 GGGAAATCATAGCAAAATGCCATGTCGTTATTTGCAGACAGAGAACCTGATATGCGCATGAG 2820
```

RESULT 2

AAS17837

ID AAS17837 standard; cdna; 3731 bp.

XX

AC AAS17837;

```
Db 2761 GGGAAATCATAGCAAAATGCCATGTCGTTATTTGCAGACAGAGAACCTGATATGCGCATGAG 2820
QY 2821 TGATCCATGCAAAAACCTTGCTGAGCTCTTTTCAATTCAGCTGSGACTTCCAAAGACTGG 2880
Db 2821 TGATCCATGCAAAAACCTTGCTGAGCTCTTTTCAATTCAGCTGSGACTTCCAAAGACTGG 2880
QY 2881 TGTTCCCGTGAATACCATCTCAGTTGGCCCTAAAGAAATACCCAGACTTCATCGATAA 2940
Db 2881 TGTTCCCGTGAATACCATCTCAGTTGGCCCTAAAGAAATACCCAGACTTCATCGATAA 2940
QY 2941 GCGGCACAGACACAGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTCAGGAAAGTGAA 3000
Db 2941 GCGGCACAGACACAGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTCAGGAAAGTGAA 3000
QY 3001 GGACAAAGCACCTCAGGCTAGCTCTATCCGACCTTCAACAGAGATGTTGCAAGAGATC 3060
Db 3001 GGACAAAGCACCTCAGGCTAGCTCTATCCGACCTTCAACAGAGATGTTGCAAGAGATC 3060
QY 3061 ATATGATGCTGATATGGAAGTTGATGGATTGAAGATTACATTGACGAAGCTTTTGACTA 3120
Db 3061 ATATGATGCTGATATGGAAGTTGATGGATTGAAGATTACATTGACGAAGCTTTTGACTA 3120
QY 3121 CAAAACCTGAATATGACAAACAGCTGGGTAATTAATGGACTACTATGGCATATAAACAGA 3180
Db 3121 CAAAACCTGAATATGACAAACAGCTGGGTAATTAATGGACTACTATGGCATATAAACAGA 3180
QY 3181 GCGTGAATACCTTAGTGGTGCCATTATGAAGCATCAAAAACCTTTTGACCGCAGAAAAGA 3240
Db 3181 GCGTGAATACCTTAGTGGTGCCATTATGAAGCATCAAAAACCTTTTGACCGCAGAAAAGA 3240
QY 3241 TGCTGAGGCCAATAGTGTGCTGAGGCGCTTGAGGAAGGAGGCAAGCCCTGTTCAA 3300
Db 3241 TGCTGAGGCCAATAGTGTGCTGAGGCGCTTGAGGAAGGAGGCAAGCCCTGTTCAA 3300
QY 3301 GAGGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACACGTTACATA 3360
Db 3301 GAGGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACACGTTACATA 3360
QY 3361 TCATCCTACATATGGGGTGTCTACAACTCAGGGGTTGAAAGAGCTCATTTTCATTAGCTT 3420
Db 3361 TCATCCTACATATGGGGTGTCTACAACTCAGGGGTTGAAAGAGCTCATTTTCATTAGCTT 3420
QY 3421 TCCCTGGTGTGTTATGACCAGCTAATCCAGATTAGAGGAGCAAGACGTAACAGGCC 3480
Db 3421 TCCCTGGTGTGTTATGACCAGCTAATCCAGATTAGAGGAGCAAGACGTAACAGGCC 3480
QY 3481 AGTTCTCAACTGTCATCTCAGGGCTCACTGAGTCAACAGATTAGTGTTCGAATGAGA 3540
Db 3481 AGTTCTCAACTGTCATCTCAGGGCTCACTGAGTCAACAGATTAGTGTTCGAATGAGA 3540
QY 3541 TTCCAGTCAGCGTTAAGCTGATATATATGTAATAGGGTGTGATCATAGAAGAAC 3600
Db 3541 TTCCAGTCAGCGTTAAGCTGATATATATGTAATAGGGTGTGATCATAGAAGAAC 3600
QY 3601 TGTTATGCAATGTTGACTACCTTTTCTTTAAACTGCATGAAGCTGAACATATATGC 3660
Db 3601 TGTTATGCAATGTTGACTACCTTTTCTTTAAACTGCATGAAGCTGAACATATATGC 3660
QY 3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTAATATGTAATGTAATTTGATTTTCATC 3720
Db 3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTAATATGTAATGTAATTTGATTTTCATC 3720
QY 3721 TTTCCTCTAAA 3731
Db 3721 TTTCCTCTAAA 3731
```

XX 08-MAY-2002 (first entry)
XX Tomato RNA-directed RNA polymerase (RdRP) cDNA.
DE
XX
XX
KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection;
KW immunoprecipitation; immunolocalisation; ss; gene therapy.
XX
OS Lycopersicon esculentum.
XX
XX Key Location/Qualifiers
FH 194..3538
FT CDS
FT /*tag= a
FT /product= "RdRP protein"
FT /EC_number= "2.7.7.48"
XX
XX US2001023067-A1.
XX
XX 20-SEP-2001.
XX
XX 08-FEB-2001; 2001US-0782874.
XX
XX 05-MAR-1997; 97US-0811583.
XX
XX (WASS/) WASSENEGER M.
XX (RIED/) RIEDEL L.
XX
XX Wassenecker M, Riedel L, Schiebel W, Sanger HL;
XX WPI; 2001-595798/67.
XX P-PSDB; AAU10006.
XX
XX New nucleic acid molecule encoding a polypeptide having the enzymatic
XX activity of RNA-directed RNA polymerase, for modulating gene expression
XX and treating cancer and virus infection in human and animals
XX
XX Claim 1; Page 15-19; 34pp; English.
XX
XX This sequence represents a cDNA encoding the tomato RNA-directed RNA
XX polymerase (RdRP) protein of the invention. The invention comprises the
XX nucleic acid and protein sequences of RdRP. The protein of the invention
XX can catalyze in vitro transcription of short single stranded RNAs into
XX DNA molecules, this transcription can be either primed by RNA or DNA
XX oligonucleotides or be unprimed. The protein may have cytosolic or
XX virulic activities. The sequences of the invention may be used in gene
XX therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA
XX sequence and a template nucleic acid molecule derived from a nucleic
XX acid molecule which causes a disease are useful for treating a disease
XX caused by the undesired expression or overexpression of a nucleic acid
XX molecule in a human, rat or mouse, by administering the molecules. This
XX system can be used in the preparation of a pharmaceutical composition
XX and for inhibiting expression of any desired gene by transferring the
XX RdRP system to organisms that either lack a comparable mechanism or do
XX not sufficiently express their own RdRP. An antibody or an antagonist or
XX inhibitor to the protein are useful for inhibiting RNA directed RNA
XX synthesis and for ensuring stable heterologous, gene expression in
XX transgenic organisms. The sequence is useful for probes and/or for the
XX control of gene expression, as primers for amplification of nucleic acid
XX molecules and as tools for the detection of expression of the cDNA
XX molecules. Additionally, nucleotide and protein sequences are useful for
XX suppression of undesired gene expression in humans and animals. The RdRP
XX is useful as a therapeutic agent for the control of cancer and virus
XX infection in humans and animals and the antibody is useful for
XX immunoprecipitation or immunolocalisation of the protein, identification
XX of polypeptides interacting with it and screening expression libraries.
XX
XX Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 other;
XX
XX Query Match 100.0%; Score 3731; DB 23; Length 3731;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATATTTCTTTACTTACTTCCACGAGGATTTGACTCATCTCCCTCCCTCAAGTCTTTTGT 60
DB 1 GAATATTTCTTTACTTACTTCCACGAGGATTTGACTCATCTCCCTCCCTCAAGTCTTTTGT 60
QY 61 GTTGATATAATAATTTGGTTGCTTCAAGTTTCAGTCACTACTGCTGGGTAGTTTTAT 120
DB 61 GTTGATATAATAATTTGGTTGCTTCAAGTTTCAGTCACTACTGCTGGGTAGTTTTAT 120
QY 121 TTTGCATAAATTCAGGGGATTTCCAGTTGGTGTAGCATTTGAAAGTCGAACATGCACTT 180
DB 121 TTTGCATAAATTCAGGGGATTTCCAGTTGGTGTAGCATTTGAAAGTCGAACATGCACTT 180
QY 181 GGAATTTGGCTACATGGGAAAGACAATTCAGTTTTCGGATTCCCTTATCTCTCTGTC 240
DB 181 GGAATTTGGCTACATGGGAAAGACAATTCAGTTTTCGGATTCCCTTATCTCTCTGTC 240
QY 241 GGAAGTGGTTAAGTCATTTCTTAGAGAAATATACAGGATATGGAATGTATGTGATGGA 300
DB 241 GGAAGTGGTTAAGTCATTTCTTAGAGAAATATACAGGATATGGAATGTATGTGATGGA 300
QY 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCATTTGCCAAAGTTCGAATTTGCCGACACAT 360
DB 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCATTTGCCAAAGTTCGAATTTGCCGACACAT 360
QY 361 AAGTCTGACAAATCATCTTTGGCTAATAACAGGCTGTATTTGGCTCTCTCTTATTT 420
DB 361 AAGTCTGACAAATCATCTTTGGCTAATAACAGGCTGTATTTGGCTCTCTCTTATTT 420
QY 421 GAAAGCTTGGGAAATGAAACTGATATTGTCCAACTGGGGCATATGTGGATCAGATGGA 480
DB 421 GAAAGCTTGGGAAATGAAACTGATATTGTCCAACTGGGGCATATGTGGATCAGATGGA 480
QY 481 TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATCACAAGTTGCAGTTGGGAAG 540
DB 481 TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATCACAAGTTGCAGTTGGGAAG 540
QY 541 TACAGAAGTTCAATTTCAATTTGGCATTTGGAATTTGGAAGAAATTTTCTTTTATCTAG 600
DB 541 TACAGAAGTTCAATTTCAATTTGGCATTTGGAATTTGGAAGAAATTTTCTTTTATCTAG 600
QY 601 TGGTTCAGTCACTATAAATTTCAAGTTTCAGTTTATATGAAAATATATGCGAGTTGCTCCA 660
DB 601 TGGTTCAGTCACTATAAATTTCAAGTTTCAGTTTATATGAAAATATATGCGAGTTGCTCCA 660
QY 661 TCGTCCATATGTGCAAAATGCTTCAGTTTCTCTCATACAGTTATTTGGTCTCTCTCGAT 720
DB 661 TCGTCCATATGTGCAAAATGCTTCAGTTTCTCTCATACAGTTATTTGGTCTCTCTCGAT 720
QY 721 CTATAGAGACTTGAATACTCTCTGTATAGCTTCTTTAAGGAACTCTCTGATGATGATG 780
DB 721 CTATAGAGACTTGAATACTCTCTGTATAGCTTCTTTAAGGAACTCTCTGATGATGATG 780
QY 781 GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGCTATCTCTAGCTTATGTTTCCA 840
DB 781 GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGCTATCTCTAGCTTATGTTTCCA 840
QY 841 GTTCGTTAGGGTGTGCTGCTTCCAAATTTCCAGGAAAGTTTTCCTACTATGCAAGACG 900
DB 841 GTTCGTTAGGGTGTGCTGCTTCCAAATTTCCAGGAAAGTTTTCCTACTATGCAAGACG 900
QY 901 TGAACAAATATTTACTTTACAGACTGGTTTTCACCTTTTTCGTTCTCTCAAAATTCGGCTCT 960
DB 901 TGAACAAATATTTACTTTACAGACTGGTTTTCACCTTTTTCGTTCTCTCAAAATTCGGCTCT 960
QY 961 GGTTCCTCAATGTCCAGCTCCGGAAGGAATTTCAATTCCTTACAGATTTTCTTCAAAAT 1020
DB 961 GGTTCCTCAATGTCCAGCTCCGGAAGGAATTTCAATTCCTTACAGATTTTCTTCAAAAT 1020
QY 1021 TAGTTCTTTTGGTACAGATGGAATGCATACCTGGCCAGCATTAATGCTACTTTTTCG 1080
DB 1021 TAGTTCTTTTGGTACAGATGGAATGCATACCTGGCCAGCATTAATGCTACTTTTTCG 1080

Db 3241 TGCTGAGCCATTAGTGTGTGAGGCGCTTGAGGAGGAGGCAAGACCTGGTTCAA 3300
QY 3301 GAGCGGTAAATCATATAGATGACATGTTACCAAGGCTTCGGCTTGTACACACTTACATA 3360
Db 3302 GAGCGGTAAATCATATAGATGACATGTTACCAAGGCTTCGGCTTGTACACACTTACATA 3360
QY 3361 TCATCCTACATATTGGGGTTGCTACAAATCAGGGGTTGAAAAGAGCTCAATTCATTAGCTT 3420
Db 3362 TCATCCTACATATTGGGGTTGCTACAAATCAGGGGTTGAAAAGAGCTCAATTCATTAGCTT 3420
QY 3421 TCCCTGTGTGTTTATACACAGCTAATCCAGATTAAGAGGACAAAGCAGCTAACAGGCC 3480
Db 3422 TCCCTGTGTGTTTATACACAGCTAATCCAGATTAAGAGGACAAAGCAGCTAACAGGCC 3480
QY 3481 AGTTCTCAACTTGTCTCTCAGGGCTCAACTGAGTACAGAGATTAGTTGAAATGAGA 3540
Db 3482 AGTTCTCAACTTGTCTCTCAGGGCTCAACTGAGTACAGAGATTAGTTGAAATGAGA 3540
QY 3541 TTCAGTCGAGCGTTAGCTGTATATATATATATATATATATATATATATATATATATAT 3600
Db 3542 TTCAGTCGAGCGTTAGCTGTATATATATATATATATATATATATATATATATATATAT 3600
QY 3601 TGTATGCAATGTTGACTACCTTTTGTCTTTAAACTGCAATGAGCTGCAACATATATGC 3660
Db 3602 TGTATGCAATGTTGACTACCTTTTGTCTTTAAACTGCAATGAGCTGCAACATATATGC 3660
QY 3661 AGTACTCTAAGAACAGATGACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 3720
Db 3662 AGTACTCTAAGAACAGATGACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 3720
QY 3721 TTTCTTCTAAA 3731
Db 3722 TTTCTTCTAAA 3731

RESULT 3

AAAS17845
ID AAAS17845 standard; cDNA; 3731 BP.

AC AAAS17845;

DI 08-MAY-2002 (first entry)

DE Tomato RNA-directed RNA polymerase (RdRP) cDNA from clone HF.

KW Tomato; RdRP: RNA-directed RNA polymerase; in vitro transcription;
cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
transgenic plant; transgenic animal; cancer; viral infection;
immunoprecipitation; immunolocalisation; RdRP/HF; mutant;
gene therapy; ss.

OS Lycopersicon esculentum.

PH Key Location/Qualifiers

FT CDS 194..3535

FT /*tag= a

FT /product= "RdRP protein"

FT /EC_number= "2.7.7.48"

FT mutation replace (31,T)

FT /*tag= b

FT mutation replace (32,G)

FT /*tag= c

FT mutation replace (2017,G)

FT /*tag= d

FT mutation replace (2264,T)

FT /*tag= e

FT mutation replace (3597,A)

FT /*tag= f

FT mutation replace (3599,A)

FT /*tag= g

PN US2001023067-A1.

PD

XX

PF

XX

XX

PR

XX

XX

PA

PA

XX

PI

XX

XX

DR

XX

PT

PT

XX

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 3731 BP; 1062 A; 672 C; 848 G; 1149 T; 0 other;

Query Match 93.7%; Score 3495; DB 23; Length 3731;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 3695; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 33 ACTCATCTACTCCCTCAAGTCTTTGTGTGTGATATAATAATTTGGTTGCTTCAGTT 92

Db 33 ACTCATCTACTCCCTCAAGTCTTTGTGTGTGATATAATAATTTGGTTGCTTCAGTT 92

QY 93 TCAGTCACACTGCTGGGTAGTTTTTTATTTTGCATACCTTCAGGGGGTATTCAGTTGGT 152

Db 93 TCAGTCACACTGCTGGGTAGTTTTTTATTTTGCATACCTTCAGGGGGTATTCAGTTGGT 152

QY 153 GTTAGCATTTGAAAGTCGAACCTGCACCTTGGAAATTTGGTACATCGGAAGCAATTCAGG 212

Db 153 GTTAGCATTTGAAAGTCGAACCTGCACCTTGGAAATTTGGTACATCGGAAGCAATTCAGG 212

QY 213 TTTTCGGATTCCCTTATCTCTCTCGGAGTGGTAACTTCTTGTAGAGAAATATA 272

Db 213 TTTTCGGATTCCCTTATCTCTCTCGGAGTGGTAACTTCTTGTAGAGAAATATA 272

QY 273 CAGGATATGGAACTGTATGTGCTATGGAGGTTAAACAGTCCCAAGGAGGATCTAGAGCAT 332

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

This sequence represents a cDNA encoding the tomato RNA-directed RNA polymerase (RdRP/HF) of the invention. The invention comprises the nucleic acid and protein sequences of RdRP. The protein of the invention can catalyze in vitro transcription of short single stranded RNAs into cDNA molecules. This transcription can be either primed by RNA or DNA oligonucleotides or be unprimed. The protein may have cytosolic or virucide activities. The sequences of the invention may be used in gene therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template nucleic acid molecule derived from a nucleic acid molecule which causes a disease are useful for treating a disease caused by the undesired expression or overexpression of a nucleic acid molecule in a human, rat or mouse, by administering the molecules. This system can be used in the preparation of a pharmaceutical composition and for inhibiting expression of any desired gene by transferring the RdRP system to organisms that either lack a comparable mechanism or do not sufficiently express their own RdRP. An antibody or an antagonist or inhibitor to the protein are useful for inhibiting RNA directed RNA synthesis and for ensuring stable heterologous, gene expression in transgenic organisms. The sequence is useful for probes and/or for the control of gene expression, as primers for amplification of nucleic acid molecules and as tools for the detection of expression of the cDNA molecules. Additionally, nucleotide and protein sequences are useful for suppression of undesired gene expression in humans and animals. The RdRP is useful as a therapeutic agent for the control of cancer and virus infection in humans and animals and the antibody is useful for immunoprecipitation or immunolocalisation of the protein, identification of polypeptides interacting with it and screening expression libraries. Note: This sequence is not shown in the specification but was created by the indexer from the wild type RdRP sequence shown in AAS17837 and the information shown in table 1.

Example 3; Page - ; 34pp; English.

New nucleic acid molecule encoding a polypeptide having the enzymatic activity of RNA-directed RNA polymerase, for modulating gene expression and treating cancer and virus infection in human and animals

Wasseneegger M, Riedel L, Schiebel W, Sanger HL;

(RIEDL) RIEDEL L.

(WASS/) WASSENEGGER M.

05-MAR-1997; 97US-0811583.

08-FEB-2001; 2001US-0782874.

20-SEP-2001.

|||||
Db 273 CAGGATATGGAACGTGATGTCATGAGGTTAAACAGTCCAAAGGAGGATCTAGAGCAT 332
QY 333 TTGCCAAGATTCAATTTGCCGACAAACATAAGTGTGCAAAATCATCACATTTGGCTTAATA 392
Db 333 TTGCCAAGATTCAATTTGCCGACAAACATAAGTGTGCAAAATCATCACATTTGGCTTAATA 392
QY 393 ACAGGCTGTATTTGGCTCTCTTATTTGAAGCTTTGGGAATGAAACATGATTTGTCC 452
Db 393 ACAGGCTGTATTTGGCTCTCTTATTTGAAGCTTTGGGAATGAAACATGATTTGTCC 452
QY 453 AACTGCGGGCATATGTGGATCAGATGGATGGCATAACTTTGAATTTCCGGATGTCAGATAT 512
Db 453 AACTGCGGGCATATGTGGATCAGATGGATGGCATAACTTTGAATTTCCGGATGTCAGATAT 512
QY 513 CAGATGACAAAGTTTGCAGTCTTTGGGAAGTACAGAAAGTTTCAATTCATTTGGCATTTGGAT 572
Db 513 CAGATGACAAAGTTTGCAGTCTTTGGGAAGTACAGAAAGTTTCAATTCATTTGGCATTTGGAT 572
QY 573 TGAAGAAATTTTCTTTTATCTAGTGGTTTCAGTGTGACTATAAAGTTCAGCTTTTCAT 632
Db 573 TGAAGAAATTTTCTTTTATCTAGTGGTTTCAGTGTGACTATAAAGTTCAGCTTTTCAT 632
QY 633 ATGAATATATGCGCAGGTTGCTCCATGCTCCATATGGTCAAAATGCTCAGTTTCTCC 692
Db 633 ATGAATATATGCGCAGGTTGCTCCATGCTCCATATGGTCAAAATGCTCAGTTTCTCC 692
QY 693 TCATACAGTTATTTGGTGTCTCCGATCTATAAGAGACTTTGAAACTCCTGTATAGCT 752
Db 693 TCATACAGTTATTTGGTGTCTCCGATCTATAAGAGACTTTGAAACTCCTGTATAGCT 752
QY 753 TCATTAAAGAACTCCTGATGATGATGAGTGGTGAAGACACAGATTTCCCTCCATCTTGA 812
Db 753 TCATTAAAGAACTCCTGATGATGATGAGTGGTGAAGACACAGATTTCCCTCCATCTTGA 812
QY 813 TAGGGTATCTTACCTATGTTTTCAGTTTCCAGTTCGTAGGGTGTTCGTCTTCCAAATTTCC 872
Db 813 TAGGGTATCTTACCTATGTTTTCAGTTTCCAGTTCGTAGGGTGTTCGTCTTCCAAATTTCC 872
QY 873 AGGAAAGTTTTTCCACTATGAGAACAGTGAACAAATATTACTTTACAGACTGGTTTCA 932
Db 873 AGGAAAGTTTTTCCACTATGAGAACAGTGAACAAATATTACTTTACAGACTGGTTTCA 932
QY 933 CCTTTTTCGTCCTCAAAATCGCTCTGGTTCCCAATGTCAGCCTCCGGAAGAAATTT 992
Db 933 CCTTTTTCGTCCTCAAAATCGCTCTGGTTCCCAATGTCAGCCTCCGGAAGAAATTT 992
QY 993 CAATTCCTACAGATTTTGTTCAAAATAGTTTCTTTGGTACAGCATGGATGCACTG 1052
Db 993 CAATTCCTACAGATTTTGTTCAAAATAGTTTCTTTGGTACAGCATGGATGCACTG 1052
QY 1053 GGCAGCATTAATGCTACTTTTCCGATTAGTTTGTCCCTCGAAGGAAATGTTGGCAT 1112
Db 1053 GGCAGCATTAATGCTACTTTTCCGATTAGTTTGTCCCTCGAAGGAAATGTTGGCAT 1112
QY 1113 GCATTGAGCATCCCTAGAGAACTGTACTATATAAAGAGTGTGTTATGATCCCGTGA 1172
Db 1113 GCATTGAGCATCCCTAGAGAACTGTACTATATAAAGAGTGTGTTATGATCCCGTGA 1172
QY 1173 GTGGCTCACTGAGCAGTATGATGGTATCTCAAGGGTAGACAACTCCAAATCTCCGT 1232
Db 1173 GTGGCTCACTGAGCAGTATGATGGTATCTCAAGGGTAGACAACTCCAAATCTCCGT 1232
QY 1233 CCATCAGTTTAGATGATGGTTTGGTGTATGTAAGAGGTCCTAGTACACCATGCAAG 1292
Db 1233 CCATCAGTTTAGATGATGGTTTGGTGTATGTAAGAGGTCCTAGTACACCATGCAAG 1292
QY 1293 TTTATTTTGGTCCAGAGGTTAATGTTTCCAAATCGGGTTCTCCGCAATTTATCTGAAG 1352
Db 1293 TTTATTTTGGTCCAGAGGTTAATGTTTCCAAATCGGGTTCTCCGCAATTTATCTGAAG 1352
QY 1353 ACATAGATAAATCTCTGCTGTTCTTTTGTGATGAGAGTGGGAGAACTGATTCCTA 1412
|||||

Db 1353 ACATAGATAAATCTCTGCTGTTCTTTGTTGATGAGGAGTGGGAGAACTGTATTCCTA 1412
QY 1413 CAGACTTATTACAAAAGCAAGTACTTGGAGTGGTGTGAGGACACACATCTATGAGAGA 1472
Db 1413 CAGACTTATTACAAAAGCAAGTACTTGGAGTGGTGTGAGGACACACATCTATGAGAGA 1472
QY 1473 TCTATCAACTCTGCGGAAAGGCTTTGTAATTTGGTGATAAAAAATTTGTAATTTCTTGCAT 1532
Db 1473 TCTATCAACTCTGCGGAAAGGCTTTGTAATTTGGTGATAAAAAATTTGTAATTTCTTGCAT 1532
QY 1533 TTTTCATCGAGCCAGTTGCGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGSCCTTA 1592
Db 1533 TTTTCATCGAGCCAGTTGCGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGSCCTTA 1592
QY 1593 CTGCAAAATGATATAAGAGCTTGGATGGGTGATTTTTCGCAGATCAAGATGTCGCAAAAT 1652
Db 1593 CTGCAAAATGATATAAGAGCTTGGATGGGTGATTTTTCGCAGATCAAGATGTCGCAAAAT 1652
QY 1653 ATGCTGCGCAGACTTGGTCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTAGGC 1712
Db 1653 ATGCTGCGCAGACTTGGTCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTAGGC 1712
QY 1713 ATGAGATTGAAGTTATCCCGATGTAAGGTTTCATGGAACCAAGCTATGCTTTTCTGATG 1772
Db 1713 ATGAGATTGAAGTTATCCCGATGTAAGGTTTCATGGAACCAAGCTATGCTTTTCTGATG 1772
QY 1773 GAATTTGGTAAATATCTGCTGACTTTGCTCATAGAGTTGCTCAAAATGTCGCTTCAAT 1832
Db 1773 GAATTTGGTAAATATCTGCTGACTTTGCTCATAGAGTTGCTCAAAATGTCGCTTCAAT 1832
QY 1833 ATACCCCATCTGCTTCCAGATTCGTTATGTTGGATATAAAGGTTGTTGGTGTGTTGATC 1892
Db 1833 ATACCCCATCTGCTTCCAGATTCGTTATGTTGGATATAAAGGTTGTTGGTGTGTTGATC 1892
QY 1893 CGGATTCATCAATGAAGTTGCTTTGGAAGAAGAGCATGCGAAATATGAATCAGACAACA 1952
Db 1893 CGGATTCATCAATGAAGTTGCTTTGGAAGAAGAGCATGCGAAATATGAATCAGACAACA 1952
QY 1953 TAAAGTTAGATGTCCTTGGATGGAGCAAAATATCAGCCTTGTATCTTAATCGTCAACTGA 2012
Db 1953 TAAAGTTAGATGTCCTTGGATGGAGCAAAATATCAGCCTTGTATCTTAATCGTCAACTGA 2012
QY 2013 TTACGCTCTTGTCTACACTTGGAGTGAAGATGAAGTTCTCGAAGAGCAAGCAAGGAAG 2072
Db 2013 TTACTCTCTTGTCTACACTTGGAGTGAAGATGAAGTTCTCGAAGAGCAAGCAAGGAAG 2072
QY 2073 CTGTAGATCAGCTTGATGCTATCTTGGATGATGATCTTTGAAGGCACAGAGGCTTTGGAAT 2132
Db 2073 CTGTAGATCAGCTTGATGCTATCTTGGATGATGATCTTTGAAGGCACAGAGGCTTTGGAAT 2132
QY 2133 TGATGTCTCTCGGAGAGAACACTAATATCTCAAGGCAATGCTAACTGTGGTTTATAAGC 2192
Db 2133 TGATGTCTCTCGGAGAGAACACTAATATCTCAAGGCAATGCTAACTGTGGTTTATAAGC 2192
QY 2193 CTGATGCTGAGCCCTTCTTCAATGATGTTGCAAACTTCCGGGCATCCAAGTTGCTCG 2252
Db 2193 CTGATGCTGAGCCCTTCTTCAATGATGTTGCAAACTTCCGGGCATCCAAGTTGCTCG 2252
QY 2253 ATTTGCGGACTAGATCAAGAAATATTTTCCAAATGGAAGAACTGATGGGATGTTGG 2312
Db 2253 ATTTGCGGACTAGATCAAGAAATATTTTCCAAATGGAAGAACTGATGGGATGTTGG 2312
QY 2313 ATGAATCCAGAACTTGGATATGTTGTCAGTGTGTTTACGTTTACGTTGCTGGAGCATG 2372
Db 2313 ATGAATCCAGAACTTGGATATGTTGTCAGTGTGTTTACGTTTACGTTGCTGGAGCATG 2372
QY 2373 GAGAGTTTCTGAGGATTTACATCCATTTAATACACAGACATCCACCAACAGTAATTTCA 2432
Db 2373 GAGAGTTTCTGAGGATTTACATCCATTTAATACACAGACATCCACCAACAGTAATTTCA 2432
QY 2433 TTCTGAAGGGAATGTTGTTTGCAGAAATCCATGCTTGCATCTCTGATATTCGTTG 2492
Db 2433 TTCTGAAGGGAATGTTGTTTGCAGAAATCCATGCTTGCATCTCTGATATTCGTTG 2492

```
QY 2493 TTTTAAAGGCTGFAAATGTTTCGAGCGCTGCACACATGCTAGATGTTGTTGATCCCTC 2552
Db 2493 TTTTAAAGGCTGFAAATGTTTCGAGCGCTGCACACATGCTAGATGTTGTTGATCCCTC 2552
QY 2553 AGAAAGGAAAGACCTCATCCGAATGAATGTTCTGGGAGTATGTTGGATGGGATATCT 2612
Db 2553 AGAAAGGAAAGACCTCATCCGAATGAATGTTCTGGGAGTATGTTGGATGGGATATCT 2612
QY 2613 ACTTTGTTGCTGGGATCAAGACATGATCCGCGCAAGCAAGTCCAGCGGATGGAATATC 2672
Db 2613 ACTTTGTTGCTGGGATCAAGACATGATCCGCGCAAGCAAGTCCAGCGGATGGAATATC 2672
QY 2673 CTCACGACCCAGCATACAGTTGGACCATGATGTCACAAATTGAGNAGTTGAAGAGTACT 2732
Db 2673 CTCACGACCCAGCATACAGTTGGACCATGATGTCACAAATTGAGNAGTTGAAGAGTACT 2732
QY 2733 TCACCAACTATATTGTAATGACAGTTTGGGAATCATAGCAAAATGCCCATGCGTATTG 2792
Db 2733 TCACCAACTATATTGTAATGACAGTTTGGGAATCATAGCAAAATGCCCATGCGTATTG 2792
QY 2793 CAGACAGAGNACCTGATATGGCCATGAGTGATCCATGCAAAAACCTTGCTGAGCTCTTTT 2852
Db 2793 CAGACAGAGNACCTGATATGGCCATGAGTGATCCATGCAAAAACCTTGCTGAGCTCTTTT 2852
QY 2853 CAATTGCAAGTGGACTTTCCAAAGACTGGTGTCCCGCTGAAATACCATCTCAGTTGCGGC 2912
Db 2853 CAATTGCAAGTGGACTTTCCAAAGACTGGTGTCCCGCTGAAATACCATCTCAGTTGCGGC 2912
QY 2913 CTAAGAATACCCAGACTTTCATGGTAAGCGGACAAAGACCAAGTATATCTCAGAAAGAG 2972
Db 2913 CTAAGAATACCCAGACTTTCATGGTAAGCGGACAAAGACCAAGTATATCTCAGAAAGAG 2972
QY 2973 TTATTGGAAGCTTTTCAGAAAGTGAAGCAAGCAAAACCTCAGGCTAGCTCTATCGCGA 3032
Db 2973 TTATTGGAAGCTTTTCAGAAAGTGAAGCAAGCAAAACCTCAGGCTAGCTCTATCGCGA 3032
QY 3033 CCTTCAAGAGAGTGTTCGAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTG 3092
Db 3033 CCTTCAAGAGAGTGTTCGAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTG 3092
QY 3093 AAGATTACATTGACGAAGCTTTGACTACAAACTGAATATGACACAAAGCTGGGTAAAT 3152
Db 3093 AAGATTACATTGACGAAGCTTTGACTACAAACTGAATATGACACAAAGCTGGGTAAAT 3152
QY 3153 TAATGGACTACTATGGCATAAACACAGAGCTGAAATACCTTAGTGGTGGCATTATGAAGG 3212
Db 3153 TAATGGACTACTATGGCATAAACACAGAGCTGAAATACCTTAGTGGTGGCATTATGAAGG 3212
QY 3213 CATCAAAAACCTTTTGACCCGAGAAAAGATGCTGAGGCCATTAGTGTGTGAGGGCCT 3272
Db 3213 CATCAAAAACCTTTTGACCCGAGAAAAGATGCTGAGGCCATTAGTGTGTGAGGGCCT 3272
QY 3273 TGAGGAGGAGGCAAGAGCTGGTTCAAGAGCGTAAATGATATAGATGACATGTTACCAA 3332
Db 3273 TGAGGAGGAGGCAAGAGCTGGTTCAAGAGCGTAAATGATATAGATGACATGTTACCAA 3332
QY 3333 AGGCTTCGGCTTGGTACCAAGCTTACATATCATCTACATATTGGGGTGGCTCAATCAGG 3392
Db 3333 AGGCTTCGGCTTGGTACCAAGCTTACATATCATCTACATATTGGGGTGGCTCAATCAGG 3392
QY 3393 GGTGAAAAGAGCTCAATTCATTAGCTTCCCTGGTGTCTTATGACAGCTAATCCAGA 3452
Db 3393 GGTGAAAAGAGCTCAATTCATTAGCTTCCCTGGTGTCTTATGACAGCTAATCCAGA 3452
QY 3453 TTAAGAGGACAAAGCACGCTAACAGGCCAGTGTCTCAACTTGTCACTCTCAGGGGCTCAAC 3512
Db 3453 TTAAGAGGACAAAGCACGCTAACAGGCCAGTGTCTCAACTTGTCACTCTCAGGGGCTCAAC 3512
QY 3513 TGAGTCACAGATTAGTGTGAAATGAGATTCAGTCGAGCGTTAAGCTGATATATATA 3572
Db 3513 TGAGTCACAGATTAGTGTGAAATGAGATTCAGTCGAGCGTTAAGCTGATATATATA 3572
```

```
QY 3573 ATGTAATAGGCTGATCATAGAAAACCTGTTATGCATTTGACTACCTTTTCTTTTA 3632
Db 3573 ATGTAATAGGCTGATCATAGAGATCTGTTATGCATTTGACTACCTTTTCTTTTA 3632
QY 3633 AAACCTGCATGAAGCTGCAACATATATCAGTACTCTAAGAAACAGATGTACAGCTAAGTA 3692
Db 3633 AAACCTGCATGAAGCTGCAACATATATCAGTACTCTAAGAAACAGATGTACAGCTAAGTA 3692
QY 3693 CTAATATGATGTGATTTGAGTTTCACTTTCTTCTTCTTAA 3731
Db 3693 CTAATATGATGTGATTTGAGTTTCACTTTCTTCTTCTTAA 3731

RESULT 4
AAT18078/c
ID AAT18078 standard; cDNA; 906 BP.
XX AAT18078;
XX 17-AUG-1996 (first entry)
XX Potato calmodulin gene.
DE Potato; calmodulin; gene expression; shoot; tuber; growth;
KW development; antisense; increase; decrease; ds.
XX Solanum tuberosum (Isolate p-PCM-1).
FH Key Location/Qualifiers
FT CDS 81..530
FT FT /*tag= a
FT FT /product= Calmodulin.
XX US5498533-A.
XX 12-MAR-1996.
XX 30-JUL-1993; 93US-0100874.
XX 30-JUL-1993; 93US-0100874.
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX An GH, Han T, Poovalah BW, Takezawa D;
XX WPI; 1996-159687/16.
XX P-PSDB; AAR81571.
XX Transgenic potato plants with increased stem and tuber elongation -
XX contain nucleic acid construct pPCM-1 comprising potato calmodulin
XX coding sequence in the sense direction
XX Claim 6; Columns 7-8; 14pp; English.
XX Controlling growth and development of potato plants can be achieved
XX by increasing or decreasing the expression of a gene encoding
XX calmodulin. Transgenic potato plants carrying sense nucleic acid
XX constructs of p-PCM-1, a cDNA clone of potato calmodulin, exhibit
XX increased shoot and tuber growth yet plants carrying antisense
XX nucleic acid constructs exhibit decreased shoot and tuber growth.
XX Sequence 906 BP; 254 A; 151 C; 219 G; 282 T; 0 other;

Query Match 0.8%; Score 30; DB 17; Length 906;
Best Local Similarity 100.0%; Pred. No. 0.00031;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3687 TAAGTACTAATATGATGTTGATTTGAGTTT 3716
Db 47 TAAGTACTAATATGATGTTGATTTGAGTTT 18

RESULT 5
```

```

AAD04371/c
ID AAD04371 standard; DNA; 25 BP.
XX
AC AAD04371;
XX
DT 04-JUL-2001 (first entry)
XX
DE RACE PCR primer, GSP400 to amplify tomato RdRP cDNA.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
KW PCR primer; RACE; ss.
XX
OS Lycopersicon esculentum.
XX
FN US6218142-B1.
XX
PD 17-APR-2001.
XX
PF 05-MAR-1997; 97US-0811583.
XX
PR 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASSENEGGER M.
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR WPI; 2001-289830/30.
XX
PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding -
XX
PS Example 3; Column 20; 31pp; English.
XX
CC The present sequence is RACE (rapid amplification of 5'-cDNA ends) PCR
CC gene specific reverse (GSP) primer, GSP400, used to amplify a cDNA
CC encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)
CC activity. This protein is capable of RNA-directed RNA synthesis, thus
CC using RNA as a template for synthesizing complementary RNA molecules.
CC RdRP nucleic acid is useful for modulating gene expression in plants,
CC humans and animals. This may lead to various physiological, developmental
CC and/or morphological changes. Transgenic plants containing RdRP nucleic
CC acid is especially useful in plant cell or tissue cultures and in plant
CC breeding. RdRP is useful in gene therapy, particularly for treating a
CC disease that is caused by the undesirable expression or overexpression of
CC a gene.
XX
SQ Sequence 25 BP; 10 A; 4 C; 7 G; 4 T; 0 other;

Query Match 0.7%; Score 25; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1838 CCATCTGCTTCCAGATTCTGTTATG 1862
Db 25 CCATCTGCTTCCAGATTCTGTTATG 1

RESULT 6
AAD04373/c
ID AAD04373 standard; DNA; 25 BP.
XX
AC AAD04373;
XX
DT 04-JUL-2001 (first entry)
XX
DE RACE PCR primer, GSP1200 to amplify tomato RdRP cDNA.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;

```

```

KW PCR primer; RACE; ss.
XX
OS Lycopersicon esculentum.
XX
PN US6218142-B1.
XX
PD 17-APR-2001.
XX
PF 05-MAR-1997; 97US-0811583.
XX
PR 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASSENEGGER M.
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR WPI; 2001-289830/30.
XX
PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding -
XX
PS Example 3; Column 20; 31pp; English.
XX
CC The present sequence is RACE (rapid amplification of 5'-cDNA ends) PCR
CC gene specific reverse (GSP) primer, GSP1200, used to amplify a cDNA
CC encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)
CC activity. This protein is capable of RNA-directed RNA synthesis, thus
CC using RNA as a template for synthesizing complementary RNA molecules.
CC RdRP nucleic acid is useful for modulating gene expression in plants,
CC humans and animals. This may lead to various physiological, developmental
CC and/or morphological changes. Transgenic plants containing RdRP nucleic
CC acid is especially useful in plant cell or tissue cultures and in plant
CC breeding. RdRP is useful in gene therapy, particularly for treating a
CC disease that is caused by the undesirable expression or overexpression of
CC a gene.
XX
SQ Sequence 25 BP; 4 A; 5 C; 10 G; 6 T; 0 other;

Query Match 0.7%; Score 25; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GCCGATGGAAATATCTCTCCAGCACCC 2683
Db 25 GCCGATGGAAATATCTCTCCAGCACCC 1

RESULT 7
AAS17838/c
ID AAS17838 standard; DNA; 25 BP.
XX
AC AAS17838;
XX
DT 08-MAY-2002 (first entry)
XX
DE Tomato RNA-directed RNA polymerase (RdRP) GSP400 PCR primer.
XX
KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection; RACE;
KW immunoprecipitation; immunolocalisation; PCR primer; GSP400;
KW gene therapy; ss.
XX
OS Lycopersicon esculentum.
XX
PN US2001023067-A1.
XX
PD 20-SEP-2001.
XX
PF 08-FEB-2001; 2001US-0782874.

```

XX PR 05-MAR-1997; 97US-0811583.
XX PA (WASS/) WASSENEGGER M.
XX PA (RIED/) RIEDEL L.
XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX DR WPI; 2001-595798/67.
XX PT New nucleic acid molecule encoding a polypeptide having the enzymatic
XX PT activity of RNA-directed RNA polymerase, for modulating gene expression
XX PT and treating cancer and virus infection in human and animals
XX PS Example 3; Page 11; 34pp; English.
XX CC This sequence represents the tomato RNA-directed RNA polymerase (RdRp)
XX CC specific PCR primer GSP400 used to obtain the 5' end of the cDNA
XX CC molecule using the rapid amplification of cDNA ends (RACE) technique.
XX CC The invention comprises the nucleic acid and protein sequences of RdRp,
XX CC the protein of the invention can catalyse in vitro transcription of
XX CC short single stranded RNAs into DNA molecules, this transcription can be
XX CC either primed by RNA or DNA oligonucleotides or be unprimed. The protein
XX CC may have cytosolic or virucide activities. The sequences of the
XX CC synthesis inhibitor. The RdRp cDNA sequence and a template nucleic acid
XX CC molecule derived from a nucleic acid molecule which causes a disease are
XX CC useful for treating a disease caused by the undesired expression or
XX CC overexpression of a nucleic acid molecule in a human, rat or mouse, by
XX CC administering the molecules. This system can be used in the preparation
XX CC of a pharmaceutical composition and for inhibiting expression of any
XX CC desired gene by transferring the RdRp system to organisms that either
XX CC lack a comparable mechanism or do not sufficiently express their own
XX CC RdRp. An antibody or an antagonist or inhibitor to the protein are
XX CC useful for inhibiting RNA directed RNA synthesis and for ensuring stable
XX CC heterologous, gene expression in transgenic organisms. The sequence is
XX CC useful for probes and/or for the control of gene expression, as primers
XX CC for amplification of nucleic acid molecules and as tools for the
XX CC detection of expression of the cDNA molecules. Additionally, nucleotide
XX CC and protein sequences are useful for suppression of undesired gene
XX CC expression in humans and animals. The RdRp is useful as a therapeutic
XX CC agent for the control of cancer and virus infection in humans and
XX CC animals and the antibody is useful for immunoprecipitation or
XX CC immunolocalisation of the protein, identification of polypeptides
XX CC interacting with it and screening expression libraries.
XX SQ Sequence 25 BP; 10 A; 4 C; 7 G; 4 T; 0 other;
Query Match 0.7%; Score 25; DB 23; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1838 CCATCTGCTTCCAGATTCGTTATG 1862
DB 25 CCATCTGCTTCCAGATTCGTTATG 1
RESULT 8
AAS17840/c
ID AAS17840 standard; DNA; 25 BP.
XX AC AAS17840;
XX DT 08-MAY-2002 (first entry)
XX TT Tomato RNA-directed RNA polymerase (RdRp) GSP1200 PCR primer.
DE DE
XX KW Tomato; RdRp; RNA-directed RNA polymerase; in vitro transcription;
KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection; RACE;
KW immunoprecipitation; immunolocalisation; PCR primer; GSP1200;
KW gene therapy; ss.
XX

OS Lycopersicon esculentum.
XX US2001023067-A1.
XX PN 20-SEP-2001.
XX PD 08-FEB-2001; 2001US-0782874.
XX PF 05-MAR-1997; 97US-0811583.
XX PR (WASS/) WASSENEGGER M.
XX XX (RIED/) RIEDEL L.
XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX DR WPI; 2001-595798/67.
XX PT New nucleic acid molecule encoding a polypeptide having the enzymatic
XX PT activity of RNA-directed RNA polymerase, for modulating gene expression
XX PT and treating cancer and virus infection in human and animals
XX PS Example 3; Page 11; 34pp; English.
XX CC This sequence represents the tomato RNA-directed RNA polymerase (RdRp)
XX CC specific PCR primer GSP1200 used to obtain the 5' end of the cDNA
XX CC molecule using the rapid amplification of cDNA ends (RACE) technique.
XX CC The invention comprises the nucleic acid and protein sequences of RdRp,
XX CC the protein of the invention can catalyse in vitro transcription of
XX CC short single stranded RNAs into DNA molecules, this transcription can be
XX CC either primed by RNA or DNA oligonucleotides or be unprimed. The protein
XX CC may have cytosolic or virucide activities. The sequences of the
XX CC invention may be used in gene therapy or as an RNA directed RNA
XX CC synthesis inhibitor. The RdRp cDNA sequence and a template nucleic acid
XX CC molecule derived from a nucleic acid molecule which causes a disease are
XX CC useful for treating a disease caused by the undesired expression or
XX CC overexpression of a nucleic acid molecule in a human, rat or mouse, by
XX CC administering the molecules. This system can be used in the preparation
XX CC of a pharmaceutical composition and for inhibiting expression of any
XX CC desired gene by transferring the RdRp system to organisms that either
XX CC lack a comparable mechanism or do not sufficiently express their own
XX CC RdRp. An antibody or an antagonist or inhibitor to the protein are
XX CC useful for inhibiting RNA directed RNA synthesis and for ensuring stable
XX CC heterologous, gene expression in transgenic organisms. The sequence is
XX CC useful for probes and/or for the control of gene expression, as primers
XX CC for amplification of nucleic acid molecules and as tools for the
XX CC detection of expression of the cDNA molecules. Additionally, nucleotide
XX CC and protein sequences are useful for suppression of undesired gene
XX CC expression in humans and animals. The RdRp is useful as a therapeutic
XX CC agent for the control of cancer and virus infection in humans and
XX CC animals and the antibody is useful for immunoprecipitation or
XX CC immunolocalisation of the protein, identification of polypeptides
XX CC interacting with it and screening expression libraries.
XX SQ Sequence 25 BP; 4 A; 5 C; 10 G; 6 T; 0 other;
Query Match 0.7%; Score 25; DB 23; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2659 GCCGATGGAATATCTCCAGCACCC 2683
DB 25 GCCGATGGAATATCTCCAGCACCC 1
RESULT 9
AAD04372/c
ID AAD04372 standard; DNA; 25 BP.
XX AC AAD04372;
XX AC AAD04372;
XX DT 04-JUL-2001 (first entry)
XX DE RACE PCR primer, GSP420 to amplify tomato RdRp cDNA.

XX Tomato: gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
 KW PCR primer; RACE; ss.
 XX
 OS Lycopersicon esculentum.
 XX
 PN US6218142-B1.
 XX
 PD 17-APR-2001.
 XX
 XX 05-MAR-1997; 97US-0811583.
 PF
 XX 05-MAR-1997; 97US-0811583.
 PR
 XX (WASS/) WASSENEGGER M.
 PA (RIED/) RIEDEL L.
 XX
 PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
 XX WPI; 2001-289830/30.
 DR
 XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
 XX polymerase enzymatic activity, useful in modulating gene expression in
 PT plants, humans and animals, as well as in plant cell/tissue cultures or
 PT plant breeding -
 PT
 XX
 PS Example 3; Column 20; 31pp; English.
 XX
 CC The present sequence is RACE (rapid amplification of 5'-cDNA ends) PCR
 CC gene specific reverse (GSP) primer, GSP420, used to amplify a cDNA
 CC encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)
 CC activity. This protein is capable of RNA-directed RNA synthesis, thus
 CC using RNA as a template for synthesizing complementary RNA molecules.
 CC RdRP nucleic acid is useful for modulating gene expression in plants,
 CC humans and animals. This may lead to various physiological, developmental
 CC and/or morphological changes. Transgenic plants containing RdRP nucleic
 CC acid is especially useful in plant cell or tissue cultures and in plant
 CC breeding. RdRP is useful in gene therapy, particularly for treating a
 CC disease that is caused by the undesirable expression or overexpression of
 CC a gene.
 XX
 SQ Sequence 25 BP; 9 A; 9 C; 4 G; 3 T; 0 other;
 XX
 Query Match 0.6%; Score 24; DB 22; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1879 TGTGGGTGTGATCCGGATTATC 1902
 DB 24 TGTGGGTGTGATCCGGATTATC 1
 XX
 RESULT 10
 AAS17839/c
 ID AAS17839 standard; DNA; 25 BP.
 XX
 AC AAS17839;
 XX
 XX 08-MAY-2002 (first entry)
 DT
 XX
 DE Tomato RNA-directed RNA polymerase (RdRP) GSP420 PCR primer.
 XX
 KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
 KW transgenic plant; transgenic animal; cancer; viral infection; RACE;
 KW immunoprecipitation; immunolocalisation; PCR primer; GSP420;
 KW gene therapy; ss.
 XX
 XX Lycopersicon esculentum.
 OS
 XX US2001023067-A1.
 PN
 XX

PD 20-SEP-2001.
 XX
 PF 08-FEB-2001; 2001US-0782874.
 XX
 PR 05-MAR-1997; 97US-0811583.
 XX
 PA (WASS/) WASSENEGGER M.
 PA (RIED/) RIEDEL L.
 XX
 PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
 XX WPI; 2001-595798/67.
 DR
 XX New nucleic acid molecule encoding a polypeptide having the enzymatic
 PT activity of RNA-directed RNA polymerase, for modulating gene expression
 PT and treating cancer and virus infection in human and animals -
 PT
 XX Example 3; Page 11; 34pp; English.
 PS
 XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
 CC specific PCR primer GSP420 used to obtain the 5' end of the cDNA
 CC molecule using the rapid amplification of cDNA ends (RACE) technique.
 CC The invention comprises the nucleic acid and protein sequences of RdRP,
 CC the protein of the invention can catalyze in vitro transcription of
 CC short single stranded RNAs into DNA molecules, this transcription can be
 CC either primed by RNA or DNA oligonucleotides or be unprimed. The protein
 CC may have cytostatic or virucide activities. The sequences of the
 CC invention may be used in gene therapy or as an RNA directed RNA
 CC synthesis inhibitor. The RdRP cDNA sequence and a template nucleic acid
 CC molecule derived from a nucleic acid molecule which causes a disease are
 CC useful for treating a disease caused by the undesired expression or
 CC overexpression of a nucleic acid molecule in a human, rat or mouse, by
 CC administering the molecules. This system can be used in the preparation
 CC of a pharmaceutical composition and for inhibiting expression of any
 CC desired gene by transferring the RdRP system to organisms that either
 CC lack a comparable mechanism or do not sufficiently express their own
 CC RdRP. An antibody or an antagonist or inhibitor to the protein are
 CC useful for inhibiting RNA directed RNA synthesis and for ensuring stable
 CC heterologous, gene expression in transgenic organisms. The sequence is
 CC useful for probes and/or for the control of gene expression, as primers
 CC for amplification of nucleic acid molecules and as tools for the
 CC detection of expression of the cDNA molecules. Additionally, nucleotide
 CC and protein sequences are useful for suppression of undesired gene
 CC expression in humans and animals. The RdRP is useful as a therapeutic
 CC agent for the control of cancer and virus infection in humans and
 CC animals and the antibody is useful for immunoprecipitation or
 CC immunolocalisation of the protein, identification of polypeptides
 CC interacting with it and screening expression libraries.
 XX
 SQ Sequence 25 BP; 9 A; 9 C; 4 G; 3 T; 0 other;
 XX
 Query Match 0.6%; Score 24; DB 23; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1879 TGTGGGTGTGATCCGGATTATC 1902
 DB 24 TGTGGGTGTGATCCGGATTATC 1
 XX
 RESULT 11
 ABL64204/c
 ID ABL64204 standard; DNA; 407 BP.
 XX
 AC ABL64204;
 XX
 XX 15-MAY-2002 (first entry)
 DT
 XX
 DE Stomach cancer related gene sequence SEQ ID NO:2541.
 XX
 KW Human; cancer; colon; breast; ovary; cesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW

gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 18-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 20-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 2541; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 847 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic

activity and can be used in gene therapy. M1 can be used for screening
an anti-neoplastic agent, and can be used for producing a product which
is the data collected with respect to the anti-neoplastic agent as a
result of M1, and the data is sufficient to convey the chemical
structure and/or properties of the agent. M1 can be used in the
treatment of cancer such as colon, breast, stomach, lung, thyroid,
oesophageal, ovarian, kidney, prostate or pancreatic cancer,
adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
carcinoma, papillary carcinoma and Wilms' tumour.
XX
SQ Sequence 407 BP; 120 A; 98 C; 92 G; 97 T; 0 other;
Query Match 0.5%; Score 20; DB 24; Length 407;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1916 TTGAGAAAGAGCATGTCGAA 1935
|||||
DB 384 TTGAGAAAGAGCATGTCGAA 365
RESULT 12
ABQ55295
ID ABQ55295 standard; cDNA; 786 BP.
XX
AC ABQ55295;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HCOB20 cDNA, SEQ ID NO:1175.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200200677-A1.
PN
XX 03-JAN-2002.
PD
XX 07-JUN-2001; 2001WO-US18569.
PF
XX 07-JUN-2000; 2000US-209467P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Birse CE, Rosen CA;
XX
XX WPI; 2002-147878/19.
DR
XX P-PSDB; ABP42218.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
XX Claim 1; SEQ ID NO 1175; 2922pp; English.
XX
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use

of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), immunodeficiencies, autoimmune oophoritis, cardiovascular disorders, respiratory disorders (e.g., anaemia), gastrointestinal disorders, and blood-related disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents cDNA encoding a human ovarian antigen of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 786 BP; 184 A; 198 C; 190 G; 198 T; 16 other;

Query Match 0.5%; Score 20; DB 24; Length 786;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935
|||||
DB 198 TTGAGAAAGAGCATGTCGAA 217

RESULT 13

AAD38196/c
ID AAD38196 standard; DNA; 1000 BP.

XX AC AAD38196;

XX DT 10-SEP-2002 (first entry)

XX DE Non-typeable Haemophilus influenzae strain 3224A BASB205 DNA #7.

XX KW BASB205 protein; infection; otitis media; pneumonia; inflammation;
XX KW invasive disease; hearing loss; fluid accumulation; nerve damage;
XX KW vaccine; sinusitis; ds.

XX OS Haemophilus influenzae.

XX PN WO200230971-A2.

XX PD 18-APR-2002.

XX PF 05-OCT-2001; 2001WO-EP11560.

XX PR 13-OCT-2000; 2000GB-0025171.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Thonnard J;

XX DR WPI; 2002-452340/48.

XX PT Novel isolated non-typeable Haemophilus influenzae BASB205 polypeptides
XX PT useful as components of vaccines useful for treating bacterial
XX PT infection such as otitis media, delayed speech learning, inflammation
XX PT of middle ear

XX PS Disclosure; Page 80; 92pp; English.

CC The invention relates to non-typeable Haemophilus influenzae BASB205
CC polypeptides and polynucleotides. Sequences of the invention are used
CC to treat non-typeable H. influenzae infections. They are also used
CC as components of vaccines useful for treating infections preferably
CC bacterial infections such as otitis media, pneumonia, sinusitis,
CC nosocomial infections, invasive diseases, chronic otitis media with
CC hearing loss, fluid accumulation in the middle ear, auditive nerve
CC damage, delayed speech learning, infections of the upper respiratory
CC tract and inflammation of the middle ear. The present sequence is
CC non-typeable H. influenzae strain 3224A BASB205 DNA upstream region.

XX SQ Sequence 1000 BP; 288 A; 151 C; 256 G; 305 T; 0 other;

Query Match 0.5%; Score 20; DB 24; Length 1000;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2414 TCCACCAACAGTAATTCAT 2433
|||||
DB 879 TCCACCAACAGTAATTCAT 860

RESULT 14

AAZ52915
ID AAZ52915 standard; cDNA; 1509 BP.

XX AC AAZ52915;

XX DT 14-MAR-2000 (first entry)

XX DE Human prostate tumor cDNA library derived EST fragment #58.

XX KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
XX KW treatment; ds.

XX OS Homo sapiens.

XX PN DEL9820190-A1.

XX PD 04-NOV-1999.

XX PF 28-APR-1998; 98DE-1020190.

XX PR 28-APR-1998; 98DE-1020190.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX DR WPI; 1999-621385/54.

XX DR P-PSDB; AAY73986, AAY73987, AAY73988.

XX PT New human nucleic acid sequences from pancreatic tumors, and related
XX PT proteins

XX PS Claim 2; Page 229-230; 502pp; German.

XX CC This invention describes novel polypeptides and their encoding nucleic
XX CC acids derived from human pancreatic tumor tissue which have cytostatic
XX CC activity. The sequences are also useful in producing pharmaceutical
XX CC compositions for treatment of pancreatic tumors. AAZ52958-253014
XX CC represent expressed sequence tag (EST) fragments derived from a human
XX CC pancreatic tumor cDNA library and which encode the proteins represented
XX CC in AAY73814-Y74252.

XX SQ Sequence 1509 BP; 416 A; 338 C; 416 G; 339 T; 0 other;

Query Match 0.5%; Score 20; DB 20; Length 1509;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935
|||||

Db 1187 TTGAGAAAGACGATGCGAA 1206

RESULT 15

AA18199

ID AAX18199 standard; cDNA to mRNA; 1730 BP.

XX AC AAX18199;

XX DT 10-MAY-1999 (first entry)

XX DE TIP49 coding sequence #2.

XX KW TIP49; TATA binding protein; TBP binding protein;

XX KW DNA transcription control; ss.

XX OS Rattus sp.

XX PN W09900419-A1.

XX PD 07-JAN-1999.

XX PF 25-JUN-1998; 98WO-JP02836.

XX PR 27-JUN-1997; 97JP-0187398.

XX PA (SUME) SUMITOMO ELECTRIC IND CO.

XX PI Kishimoto T, Makino Y, Niwa S, Tamura T;

XX DR WPI; 1999-095682/08.

XX DR P-PSDB; AAW74417.

XX TT TBP-binding protein with DNA helicase and ATPase activities - gene

XX PT encoding it, and antibodies recognising it.

XX PS Claim 10; Page 43-45; 64pp; Japanese.

XX CC This sequence encodes the TATA-binding protein (TBP) binding protein,

CC designated TIP49, of the invention. TIP49 and its associated nucleic

CC acids and antibodies are useful in investigation of the process of DNA

CC transcription control by TBP in vivo, and in detection of the blocking of

CC DNA transcription.

XX SQ Sequence 1730 BP; 469 A; 401 C; 476 G; 384 T; 0 other;

Query Match 0.5%; Score 20; DB 20; Length 1730;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGACGATGCGAA 1935

Db 1333 TTGAGAAAGACGATGCGAA 1352

RESULT 16

AAC59220

ID AAC59220 standard; cDNA; 1744 BP.

XX AC AAC59220;

XX DT 29-JAN-2001 (first entry)

XX DE Human secreted protein cDNA sequence #6.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;

XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;

XX KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;

XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

XX KW neurological disease; infection; human; secreted protein; ss.

XX OS Homo sapiens.

XX

PN W0200055199-A1.

XX 21-SEP-2000.

XX 09-MAR-2000; 2000WO-US06014.

XX 12-MAR-1999; 99US-0124095.

XX PR 11-JUN-1999; 99US-0138598.

XX PR 03-DEC-1999; 99US-0168665.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM, Komatsoulis G;

XX DR WPI; 2000-572359/53.

XX DR P-PSDB; AAB27799.

XX PT Isolated nucleic acid molecule encoding a human secreted protein is

XX TT used in preventing, treating or ameliorating a medical condition -

XX PS Claim 1; Page 343; 433pp; English.

XX CC The invention relate to the isolation of genes AAC59215-C59261 encoding

CC 47 human secreted proteins AAB27794-B27840. The genes can be used to

CC generate fusion proteins by linking to the gene for the human

CC immunoglobulin G Fc portion for increasing the stability of

CC the fusion protein as compared to the human protein only. The genes and

CC proteins are useful for preventing, ameliorating or treating medical

CC conditions, e.g. by protein or gene therapy. The genes are isolated

CC from a range of human tissues disclosed in the specification. The

CC nucleic acids, proteins, antibodies and (ant)agonists are useful in

CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast

CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone

CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune

CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative

CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)

CC wound healing; (e) neurological diseases e.g. cerebral anoxia and

CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.

XX SQ Sequence 1744 BP; 476 A; 403 C; 480 G; 384 T; 1 other;

Query Match 0.5%; Score 20; DB 21; Length 1744;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGACGATGCGAA 1935

Db 1342 TTGAGAAAGACGATGCGAA 1361

RESULT 17

AAX24059

ID AAX24059 standard; DNA; 1750 BP.

XX AC AAX24059;

XX DT 30-JUN-1999 (first entry)

XX DE Human regulatory molecule HRM-1 DNA.

XX KW Human regulatory molecule; HRM-1; cytostatic activity; immune modulator;

XX KW transcription factor; enhancer; cell proliferation stimulation; cancer;

XX KW treatment; microarray; detection; diagnosis; cell proliferation disease;

XX KW leukemia; lymphoma; myeloma; adenocarcinoma; sarcoma; bladder; bone;

XX KW brain; lung; liver; ovary; skin; teratocarcinoma; immune response;

XX KW allergy; asthma; diabetes; multiple sclerosis; Grave's disease;

XX KW myasthenia gravis; ss.

XX OS Homo sapiens.

XX

```

PN WO9915658-A2.
PD
PP 01-APR-1999.
PP
PP 22-SEP-1998; 98WO-US19839.
PP
PP 23-SEP-1997; 97US-0933750.
PP
PP (INCY-) INCYTE PHARM INC.
PP
PP Au-Young J, Bandman O, Guegler KJ, Hillman JL, Lal P;
PI Shah P, Yue H;
PI
XX WPI; 1999-254710/21.
XX
XX P-PSDB; AAW93945.
XX
XX New human regulatory molecules
XX
XX Claim 5; Page 70-71; 76pp; English.
XX
XX This invention describes novel human regulatory molecules (HRM) which
CC have cytostatic activity and act as immune modulators, transcription
CC factors or enhancers. The HRMs can be used to stimulate cell
CC proliferation. Antagonists and agonists of the proteins of the invention
CC can be used to treat cancer. The encoding nucleic acids can be used in
CC microarrays to detect polynucleotides (and their expression levels) that
CC encode HRMs in a biological sample. The HRMs and microarrays can be used
CC to diagnose, treat or prevent cell proliferation diseases especially cancer,
CC e.g. leukemia, lymphoma, myeloma, adenocarcinoma, sarcoma, cancer of e.g.
CC bladder, bone, brain, lung, liver, ovary, skin, etc, teratocarcinoma, or
CC to treat or prevent immune responses e.g. allergies, asthma, diabetes,
CC multiple sclerosis, Grave's disease or myasthenia gravis.
XX
XX Sequence 1750 BP; 472 A; 406 C; 485 G; 387 T; 0 other;
SQ
Query Match 0.5%; Score 20; DB 20; Length 1750;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935
Db 1355 TTGAGAAAGAGCATGTCGAA 1374

RESULT 18
ABA92477
ID ABA92477 standard; cDNA; 1750 BP.
XX
XX ABA92477;
XX
XX 18-MAR-2002 (first entry)
XX
XX Human TIP49 encoding cDNA.
XX
XX Human; STAP1; Skp2; cyclin A-interacting protein; TIP48; TIP49;
XX Skp2-associated protein one; anti-cancer; anti-proliferative; cancer;
XX cytostatic; transcription regulation; non-cMyc expressing cancer; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 77..1447
XX /*tag= a
XX /*product= "TIP49"
XX
XX WO200185980-A2.
XX
XX 15-NOV-2001.
XX
XX 11-MAY-2001; 2001WO-EP05404.
XX
XX 12-MAY-2000; 2000GB-0011439.
XX
XX

PA (NOVS ) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.
XX
XX Gstaiger MGC, Krek W;
XX
XX WPI; 2002-049448/06.
XX
XX P-PSDB; ABB04982.
XX
XX Complex containing a TIP48 or a TIP49 family member and various other
XX proteins is useful to screen for new agents to treat cancers that do
XX not express cMyc .
XX
XX Example 8; Fig 4; 34pp; English.
XX
XX The present invention describes a complex (I) comprising a TIP49 family
XX member and STAP1 (Skp1(a cyclin A-interacting protein)-associated
XX protein one), pifoldin, RPB5 and/or RNP1, preferably in a 1:1:1:1
XX ratio. Also described are: (1) a transcription regulatory complex
XX comprising TIP48 and/or TIP49 and three or more proteins or
XX polypeptides; (2) identifying an agent active against cancer cells that
XX do not express c-Myc, comprising contacting a member of the TIP49 family
XX with a test compound and measuring TIP49 enzymatic or ligand binding
XX activity; and (3) an anti-cancer agent, preferably an anti-proliferative
XX agent, particularly a nucleic acid or antibody identified by the method
XX of (2). (1) has cytostatic activity, and can be used in transcription
XX regulation. (I) can be used for screening new agents for the treatment
XX of non-cMyc expressing cancers. The present sequence encodes human
XX TIP49 from the present invention.
XX
XX Sequence 1750 BP; 476 A; 403 C; 484 G; 387 T; 0 other;
SQ
Query Match 0.5%; Score 20; DB 24; Length 1750;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935
Db 1350 TTGAGAAAGAGCATGTCGAA 1369

RESULT 19
AAK72348
ID AAK72348 standard; DNA; 38653 BP.
XX
XX AAK72348;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27160.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 18-APR-2000; 2000US-0190076.
XX
XX 19-MAY-2000; 2000US-0198123.
XX
XX 07-JUN-2000; 2000US-0205515.
XX
XX 28-JUN-2000; 2000US-0209467.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.

```

PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0225279.
PR 22-AUG-2000; 2000US-0225681.
PR 22-AUG-2000; 2000US-0225686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232968.
PR 14-SEP-2000; 2000US-0232997.
PR 14-SEP-2000; 2000US-0232998.
PR 14-SEP-2000; 2000US-0232999.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX Disclosure; SEQ ID NO 27160; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting

CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX

SQ Sequence 38653 BP; 10472 A; 7264 C; 8237 G; 12680 T; 0 other;

Query Match 0.5%; Score 20; DB 22; Length 38653;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2612 TACTTGTGTTGGGATCA 2631
 |||||
 DB 11586 TACTTGTGTTGGGATCA 11605

RESULT 20
 AAD04374
 ID AAD04374 standard; DNA; 34 BP.

XX AAD04374;
 XX
 DT 04-JUL-2001 (first entry)
 XX
 DE Forward PCR primer, P127Bam1 to amplify tomato RdRP cDNA.

XX Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
 KW PCR primer; ss.

XX Lycopersicon esculentum.

OS

XX US6218142-B1.

XX 17-APR-2001.

XX 05-MAR-1997; 97US-0811583.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASSENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-289830/30.

XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
 PT polymerase enzymatic activity, useful in modulating gene expression in
 PT plants, humans and animals, as well as in plant cell/tissue cultures or
 PT plant breeding -

XX Example 3; Column 21; 31pp; English.

XX The present sequence is forward PCR primer, P127Bam1, used to amplify a
 CC cDNA encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)
 CC activity. This primer contains BamHI site to enhance the cloning
 CC efficiency. This protein is capable of RNA-directed RNA synthesis, thus
 CC using RNA as a template for synthesising complementary RNA molecules.
 CC RdRP nucleic acid is useful for modulating gene expression in plants,
 CC humans and animals. This may lead to various physiological, developmental
 CC and/or morphological changes. Transgenic plants containing RdRP nucleic
 CC acid is especially useful in plant cell or tissue cultures and in plant
 CC breeding. RdRP is useful in gene therapy, particularly for treating a
 CC disease that is caused by the undesirable expression or overexpression of
 CC a gene.

XX Sequence 34 BP; 8 A; 15 C; 4 G; 7 T; 0 other;

SQ Sequence 34 BP; 8 A; 15 C; 4 G; 7 T; 0 other;

Best Local Similarity 100.0%; Pred. No. 95;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ACTCATCACTCCCTCAAG 51
 |||||
 DB 16 ACTCATCACTCCCTCAAG 34

RESULT 21
 AAS17841

ID AAS17841 standard; DNA; 34 BP.

XX AAS17841;

XX 08-MAY-2002 (first entry)

XX Tomato RNA-directed RNA polymerase (RdRP) P127Bam1 PCR primer.

XX Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
 KW transgenic plant; transgenic animal; cancer; viral infection; RACE;
 KW immunoprecipitation; immunolocalisation; PCR primer; P127Bam1;
 KW gene therapy; ss.

XX Lycopersicon esculentum.

XX US2001023067-A1.

XX 20-SEP-2001.

XX 08-FEB-2001; 2001US-0782874.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASSENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-595798/67.

XX New nucleic acid molecule encoding a polypeptide having the enzymatic
 PT activity of RNA-directed RNA polymerase, for modulating gene expression
 PT and treating cancer and virus infection in human and animals -

XX Example 3; Page 12; 34pp; English.

XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
 CC specific PCR primer P127Bam1 used to clone the full length RdRP cDNA
 CC molecule of the invention. The invention comprises the nucleic acid and
 CC protein sequences of RdRP, the protein of the invention can catalyse in
 CC vitro transcription of short single stranded RNAs into DNA molecules.
 CC this transcription can be either primed by RNA or DNA oligonucleotides
 CC or be unprimed. The protein may have cytosolic or virucide activities.
 CC The sequences of the invention may be used in gene therapy or as an RNA
 CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template
 CC nucleic acid molecule derived from a nucleic acid molecule which causes
 CC a disease are useful for treating a disease caused by the undesired
 CC expression or overexpression of a nucleic acid molecule in a human, rat
 CC or mouse, by administering the molecules. This system can be used in the
 CC preparation of a pharmaceutical composition and for inhibiting
 CC expression of any desired gene by transferring the RdRP system to
 CC organisms that either lack a comparable mechanism or do not sufficiently
 CC express their own RdRP. An antibody or an antagonist or inhibitor to the
 CC protein are useful for inhibiting RNA directed RNA synthesis and for
 CC ensuring stable heterologous, gene expression in transgenic organisms.
 CC The sequence is useful for probes and/or for the control of gene
 CC expression, as primers for amplification of nucleic acid molecules and
 CC as tools for the detection of expression of the cDNA molecules.
 CC Additionally, nucleotide and protein sequences are useful for
 CC suppression of undesired gene expression in humans and animals. The RdRP
 CC is useful as a therapeutic agent for the control of cancer and virus
 CC infection in humans and animals and the antibody is useful for

CC immunoprecipitation or immunolocalisation of the protein, identification
CC of polypeptides interacting with it and screening expression libraries.

XX SQ Sequence 34 BP; 8 A; 15 C; 4 G; 7 T; 0 other;

Query Match 0.5%; Score 19; DB 23; Length 34;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ACTCATCTCCCTCAAG 51
DB 16 ACTCATCTCCCTCAAG 34

RESULT 22
AAH81903/c
ID AAH81903 standard; DNA; 255 BP.

XX AC AAH81903;

XX DT 21-SEP-2001 (first entry)

XX DE Rat differential transcription-associated cDNA SEQ ID 412.

XX KW Differential transcription; human; rat; tumour cell; cytostatic;
XX KW Ras modulator; Class II tumour suppressor gene; gene therapy; ss.

XX OS Rattus sp.

XX PN WO200157058-A2.

XX PD 09-AUG-2001.

XX PF 31-JAN-2001; 2001WO-EP01003.

XX PR 31-JAN-2000; 2000DE-1004102.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Rosenthal A, Hinzner B, Schaefer R, Zuber J, Tchernitsa O;

XX PI Grips M, Hellriegel M, Schmitz A, Sers C;

XX DR WPI; 2001-483415/52.

XX Nucleic acids differentially expressed between tumor and normal cells,
PT useful for diagnosis or therapy of tumors and for screening active
PT agents -

XX Claim 6; Page 492; 579pp; German.

XX CC This invention describes a nucleic acid (I) with differential expression
CC between tumor and normal cells and which has cytostatic activity. (I)
CC work as modulators of Ras activity by inducing expression of tumour
CC suppressor genes. (I), and polypeptides encoded by them, are useful as
CC targets for diagnosis or therapy and in screening to determine the
CC effects of an active compound (potential pharmaceutical) on a cell line,
CC particularly for diagnosis and treatment of tumors, especially by
CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
CC methods) or by modulating the amount and/or location of (I)-encoded
CC polypeptides (by administration of the polypeptide or its activator,
CC antibody (optionally as a conjugate) or inhibitor). The method allows
CC identification of many Class II tumour suppressor genes (i.e. genes that
CC are not primary targets for tumour-initiating mutations).
CC AAH81492-AAH82376 represent the human and rat derived nucleic acid
CC fragments described in the method of the invention.

XX SQ Sequence 255 BP; 98 A; 47 C; 59 G; 51 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 255;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1758 ATGCTTTTCTGATGAAT 1776

DB 113 ATGCTTTTCTGATGAAT 95

RESULT 23
ABN67264/c
ID ABN67264 standard; DNA; 327 BP.

XX AC ABN67264;

XX DT 01-JUL-2002 (first entry)

XX DE Streptococcus polynucleotide SEQ ID NO 2441.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX OS Streptococcus agalactiae.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX PI Tettelin H;

XX DR WPI; 2002-352536/38.

XX DR P-PSDB; ABP26633.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

XX Claim 7; Page 3395; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

XX SQ Sequence 327 BP; 99 A; 57 C; 74 G; 97 T; 0 other;

Query Match 0.5%; Score 19; DB 24; Length 327;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2416 CACCAACAGTAATTCATT 2434

DB 292 CACCAACAGTAATTCATT 274

```
RESULT 24
AAC31397/C
ID AAC31397 standard; cDNA; 369 BP.
XX
XX AAC31397;
AC
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein 5' EST, SEQ ID NO: 35472.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST ) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI; 2000-500381/45.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PT
XX
XX Claim 1; SEQ ID 35472; 71pp + CD-ROM; English.
PS
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
XX
XX Sequence 369 BP; 72 A; 76 C; 83 G; 138 T; 0 other;
SQ
Query Match 0.5%; Score 19; DB 21; Length 369;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2065 AAAGGAAGCTGTAGATCAG 2083
Db 62 AAAGGAAGCTGTAGATCAG 44
RESULT 25
AA187435
ID AA187435 standard; cDNA; 452 BP.
XX
XX AA187435;
AC
XX 06-NOV-2001 (first entry)
DT
XX
XX Human polynucleotide SEQ ID NO 7495.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW
```

```
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
OS
XX Homo sapiens.
XX
XX WO200164835-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001WO-US04927.
PF
XX
XX 28-FEB-2000; 2000US-0515126.
PR
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-514838/56.
DR
XX
XX P-PSDB; AAO07504.
DR
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PT
XX
XX Claim 1; SEQ ID NO 7495; 1399pp + Sequence Listing; English.
PS
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 452 BP; 213 A; 53 C; 77 G; 104 T; 5 other;
SQ
Query Match 0.5%; Score 19; DB 22; Length 452;
Best Local Similarity 100.0%; Pred. No. 1e-02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 578 AAATTTTCTTTCTTTTAT 596
Db 390 AAATTTTCTTTCTTTTAT 408
RESULT 26
AA149104/C
ID AA149104 standard; DNA; 456 BP.
XX
XX AA149104;
AC
XX
XX 23-SEP-1997 (first entry)
DT
XX
XX Partial DNA clone Acii#1-426 encoding immunostimulatory peptide.
DE
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO9700067-A1.
PN
XX
XX 03-JAN-1997.
PD
```

```

XX PF 14-JUN-1996; 95WO-US10375.
XX PR 15-JUN-1995; 95US-0000254.
XX PA (UYVI-) UNIV VICTORIA.
XX PI Nano FE;
XX DR WPI; 1997-077347/07.
XX PT New immuno-stimulatory peptide(s) of Mycobacterium tuberculosis -
XX PT useful in vaccines, diagnostic skin test, immunoassay and gene
XX PT isolation
XX PS Claim 1; Page 34; 79pp; English.
XX CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
XX CC which encode partial sequences of immunostimulatory peptides. Each of
XX CC the clones encode at least one immunostimulatory T cell epitope. The
XX CC clones were identified by testing over 300 fusion clones (alkaline
XX CC phosphatase-M. tuberculosis peptide fusions) for their ability to
XX CC stimulate interferon (IFN)-gamma production. 80 clones were initially
XX CC designated to have some ability to stimulate IFN-gamma production, of
XX CC which 76 are shown in AAT49100-175. These sequences can be used to
XX CC obtain the full length M. tuberculosis genes and corresponding proteins
XX CC using standard techniques. The peptides are useful in vaccines, as
XX CC reagents in an improved tuberculin skin test (especially using peptides
XX CC different from those used in vaccines so as to allow differentiation
XX CC between vaccinated and infected subjects) and as immunoassay reagents
XX CC for detecting specific antibodies. An advantage of these peptides is
XX CC that they stimulate production of IFN-gamma (critical for a protective
XX CC immune response to M. tuberculosis) by CD4-positive T cells. The protein
XX CC encoded by this sequence has amino acid similarity to a dipeptide
XX CC transport protein.
XX SQ Sequence 456 BP; 63 A; 121 C; 165 G; 98 T; 9 other;
      Query Match 0.5%; Score 19; DB 18; Length 456;
      Best Local Similarity 100.0%; Pred. No. 1e+02;
      Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2406 ACAGCAGATCCACCAACAG 2424
      |||||
DB 296 ACAGCAGATCCACCAACAG 278

RESULT 27
ABA59152/c
ID ABA59152 standard; DNA; 461 BP.
XX AC ABA59152;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #7457.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00569.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX

```

```

PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 1; SEQ ID NO 7457; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC fetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;
      Query Match 0.5%; Score 19; DB 22; Length 461;
      Best Local Similarity 100.0%; Pred. No. 1e+02;
      Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 629 TCATATGAAATATATGTC 647
      |||||
DB 56 TCATATGAAATATATGTC 38

RESULT 28
AAK07338/c
ID AAK07338 standard; DNA; 461 BP.
XX AC AAK07338;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 7329.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX

```

PS Example 4; SEQ ID NO: 7329; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 461;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGCG 647

DB 56 TCATATGAAATATATGCG 38

RESULT 29

AAK33106/c

ID AAK33106 standard; DNA; 461 BP.

XX AAK33106;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 7663.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 7663; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.

XX Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 461;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGCG 647

DB 56 TCATATGAAATATATGCG 38

RESULT 30

AAI38901/c

ID AAI38901 standard; DNA; 461 BP.

XX AAI38901;

XX 17-OCT-2001 (first entry)

XX Probe #7587 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488997/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 7587; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 461;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGCG 647

DB 56 TCATATGAAATATATGCG 38

RESULT 31

ABS07923/c

ID ABS07923 standard; DNA; 461 BP.

XX ABS07923;

XX 19-AUG-2002 (first entry)

XX Human genome-derived single exon probe from lung SEQ ID No 7914.

Human; ds: single exon probe; asthma; lung cancer; COPD; ILD;
chronic obstructive pulmonary disease; interstitial lung disease;
familial idiopathic pulmonary fibrosis; neurofibromatosis;
tuberculous sclerosis; Gaucher's disease; Niemann-Pick disease;
Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
primary ciliary dyskinesia; pulmonary hypertension;
hyaline membrane disease.

Homo sapiens.

W0200186003-A2.

15-NOV-2001.

30-JAN-2001; 2001WO-US00665.

04-FEB-2000; 2000US-180312P.

26-MAY-2000; 2000US-207456P.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-234687P.

27-SEP-2000; 2000US-236359P.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2002-114183/15.

Spatially-addressable set of single exon nucleic acid probes, used to
measure gene expression in human lung samples -

Claim 1; SEQ ID No 7914; 634pp; English.

The invention relates to a spatially-addressable set of single exon
nucleic acid probes for measuring gene expression in a sample derived
from human lung comprising single exon nucleic acid probes having one of
12614 nucleic acid sequences mentioned in the specification, or their
complements or the 12387 open reading frames derived from the 12614
probes. Also included are a microarray comprising the novel set of
probes; the novel set of probes which hybridise at high stringency to a
nucleic acid expressed in the human lung; measuring gene expression in a
sample derived from human lung, comprising (a) contacting the array with
a collection of detectably labeled nucleic acids derived from human lung
mRNA, and (b) measuring the label detectably bound to each probe of
the array; identifying exons in a eukaryotic genome, comprising
(a) algorithmically predicting at least one exon from genomic sequences
of the eukaryote; and (b) detecting specific hybridisation of detectably
labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
having a fragment identical to the predicted exon, the probe is included
in the above mentioned microarray; assigning exons to a single gene,
comprising (a) identifying exons from genomic sequence by the method
above and (b) measuring the expression of each of the exons in several
tissues and/or cell types using hybridisation to a single exon
microarrays having a probe with the exon, where a common pattern of
expression of the exons in the tissues and/or cell types indicates that
the exons should be assigned to a single gene; a peptide comprising one
of 12011 sequences, mentioned in the specification, or encoded by the
probes/open reading frames (ORF). The probes are used for gene
expression analysis, and for identifying exons in a gene, particularly
using human lung derived mRNA and for the study of lung diseases
such as asthma, lung cancer, chronic obstructive pulmonary disease
(COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
and hyaline membrane disease. The present sequence is a single exon
probe of the invention.

CC Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic
format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.

Query Match 0.5%; Score 19; DB 24; Length 461;

Best Local Similarity 100.0%; Pred. No. 1e-02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAATATATATGCG 647

Db 56 TCATATGAATATATGCG 38

RESULT 32

AAF22601/C

ID AAF22601 standard; cDNA; 757 BP.

XX AAF22601;

DT 26-MAR-2001 (first entry)

Human breast cancer associated antigen nucleotide sequence SEQ ID NO:180.

Human; breast cancer; gastric cancer; prostate cancer; diagnosis;

cancer associated antigen; cytostatic; cancer vaccine; ss.

OS Homo sapiens.

PN W0200073801-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14749.

PR 28-MAY-1999; 99US-0136526.

PR 10-SEP-1999; 99US-0153454.

XX (LUDW-) LUDWIG INST CANCER RES.

PI Obata Y;

DR WPI; 2001-025274/03.

Nucleic acids encoding breast, gastric and prostate cancer associated

antigen precursors, useful for diagnosing and treating a condition

characterized by expression of an abnormal amount of a protein, e.g.

cancer -

Claim 50; Page 332; 799pp; English.

AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
represent nucleotide sequences encoding human breast, gastric and
prostate cancer associated antigen precursors (CAAP) respectively.
AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
represent human breast, gastric and prostate CAAP protein sequence
respectively. CAAPs have cytostatic activity and can be used in the
production of cancer vaccines. The human CAAP proteins, peptides, nucleic
acids or anti-CAAP antibodies are useful for diagnosing and treating a
condition characterised by expression of an abnormal amount of a protein,
e.g. cancer.

Sequence 757 BP; 146 A; 232 C; 244 G; 132 T; 3 other;

Query Match 0.5%; Score 19; DB 22; Length 757;

Best Local Similarity 100.0%; Pred. No. 1e-02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2775 ATGCCCAATGTCGATTTCG 2793

Db 257 ATGCCCAATGTCGATTTCG 239

RESULT 33
AAC38539/C
ID AAC38539 standard; DNA; 783 BP.
XX AC AAC38539;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21334.
XX HYbridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PE 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130810.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

```
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.5%; Score 19; DB 21; Length 783;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2094 TCTGTCATGATCTTTGAA 2112
DB 597 TCTGTCATGATCTTTGAA 579

RESULT 34
AAI94184
ID AAI94184 standard; cDNA; 889 BP.
XX
AC AAI94184;
```

```
XX 13-NOV-2001 (first entry)
XX Human neuroblastoma expressed polynucleotide SEQ ID NO 259.
XX Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX Homo sapiens.
XX WC200166719-A1.
XX 13-SEP-2001.
XX 02-MAR-2001; 2001WO-JP01629.
XX 07-MAR-2000; 2000JP-0159195.
XX (CHIB-) CHIBA PREFECTURE.
XX (HISM) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
XX WPI; 2001-565584/63.
XX Nucleic acids originating in gene expressed in human neuroblastoma,
XX useful as probe or primer in diagnosing prognosis of human
XX neuroblastoma, malignancy and susceptibility indicator or tumour marker
XX for anti-cancer agents.
XX Claim 1; Page 230-231; 2979pp; Japanese.
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
XX human neuroblastoma. The nucleic acids are applicable as a probe or
XX primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX susceptibility indicators or tumour markers for anti-cancer agents. The
XX gene information for diagnosing prognosis is related to factors similar
XX to that for N-myc and TrkA genes.
XX Sequence 889 BP; 240 A; 157 C; 142 G; 317 T; 33 other;

Query Match 0.5%; Score 19; DB 22; Length 889;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 627 TTTCATATGAAATATATG 645
DB 533 TTTCATATGAAATATATG 551

RESULT 35
AAU05849
ID AAX05849 standard; cDNA; 1296 BP.
XX
AC AAX05849;
XX
XX 04-MAY-1999 (first entry)
XX
XX Mouse pheromone receptor VR10 coding sequence.
XX Pheromone receptor; signal transduction; fertility; behaviour;
XX reproduction; rodent; insect; mouse; ss.
XX Mus sp.
XX WO9900422-A1.
XX 07-JAN-1999.
XX 30-JUN-1998; 98WO-US13680.
XX 30-JUN-1997; 97US-0051284.
XX (HARD) HARVARD COLLEGE.
PA
```

XX PI Buck L, Dulac C, Herrada G, Matsunami H;
 XX WPI: 1999-095684/08.
 DR P-PSDB; AAW94910.
 XX
 PT New isolated pheromone receptor polypeptides - used to develop
 PT products for controlling fertility and behaviour in vertebrates and
 PT invertebrates
 XX
 PS Claim 23; Page 183; 308pp; English.
 XX
 CC The invention relates to polynucleotide sequences encoding mammalian
 CC pheromone receptor polypeptides. The polypeptides are expressed in
 CC murine and rat vomeronasal organ. The products can be used for modifying
 CC pheromone activity, e.g. for decreasing pheromone receptor mediated
 CC signal transduction. They can be used for controlling fertility and
 CC behaviour in vertebrates and invertebrates. Compositions comprising the
 CC polypeptides are particularly useful in e.g. controlling fertility in
 CC livestock and controlling reproduction in rodents or insects by
 CC interrupting the normal behaviours of rodents or insects that result in
 CC reproduction. The present sequence represents the coding sequence of
 CC mouse pheromone receptor VR10.
 XX
 SQ Sequence 1296 BP; 414 A; 229 C; 238 G; 415 T; 0 other;
 Query Match 0.5%; Score 19; DB 20; Length 1296;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2806 TGATATGCCCATGAGTGAT 2824
 Db 1152 TGATATGCCCATGAGTGAT 1170
 RESULT 36
 AAA61289
 ID AAA61289 standard; DNA; 1405 BP.
 AC AAA61289;
 DT 18-OCT-2000 (first entry)
 XX
 DE Human secreted protein gene 30 clone HYBAR26.
 XX
 KW Human; secreted protein; fusion protein; gene therapy;
 KW protein therapy; diagnosis; tissue; cancer; tumour; AIDS;
 KW autoimmune disorder; allergy; cardiovascular; viral; bacterial;
 KW fungal infection; immunosuppressive; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200029422-A1.
 XX
 PD 25-MAY-2000.
 XX
 XX 09-NOV-1999; 99WO-US26409.
 PF
 PR 12-NOV-1998; 98US-0108207.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ni J, Ruben SM, Rosen CA, Ebner R, Florence KA, Young PE;
 PI Birse CE, Carter KC, Komatsoulis G;
 XX
 XX WPI: 2000-387729/33.
 DR
 XX Novel human secreted proteins useful for diagnosing, preventing,
 PT treating and ameliorating a medical condition e.g. cardiovascular
 PT disease -
 XX
 XX Claim 1; Page 248; 295pp; English.
 PS
 XX

CC The present sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number and the clone it was derived
 CC from are given in the descriptor line.
 CC The invention relates to 31 novel genes and their fragments (nucleic
 CC acid sequences: AAA61260-A61293; amino acid sequences AAB12301-B12371)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also pathological conditions
 CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Specific uses are described for each of the 31
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC and include products for the diagnosis or treatment of cancer, tumours,
 CC AIDS, autoimmune disorders, allergy, cardiovascular disorders, viral,
 CC bacterial and fungal infection. The genes are used to generate fusion
 CC proteins by linking to the gene a human immunoglobulin portion (AAA61251)
 CC for increasing stability of the fused protein as compared to the
 CC secreted protein only.
 XX
 SQ Sequence 1405 BP; 410 A; 264 C; 312 G; 412 T; 7 other;
 Query Match 0.5%; Score 19; DB 21; Length 1405;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2300 ATGGGATGTTGGATGAAT 2318
 Db 822 ATGGGATGTTGGATGAAT 840

RESULT 37
 AAX05850
 ID AAX05850 standard; cDNA; 1521 BP.
 XX
 AC AAX05850;
 DT 04-MAY-1999 (first entry)
 XX
 DE Mouse pheromone receptor VR11 coding sequence.
 XX
 KW Pheromone receptor; signal transduction; fertility; behaviour;
 KW reproduction; rodent; insect; mouse; ss.
 XX
 OS Mus sp.
 XX
 PN WO9900422-A1.
 XX
 PD 07-JAN-1999.
 XX
 PF 30-JUN-1998; 98WO-US13680.
 XX
 PR 30-JUN-1997; 97US-0051284.
 XX
 PA (HARD) HARVARD COLLEGE.
 XX
 PI Buck L, Dulac C, Herrada G, Matsunami H;
 XX
 DR WPI: 1999-095684/08.
 DR P-PSDB; AAW94911.
 XX
 PT New isolated pheromone receptor polypeptides - used to develop
 PT products for controlling fertility and behaviour in vertebrates and
 PT invertebrates
 XX
 PS Claim 23; Page 184; 308pp; English.
 XX
 CC The invention relates to polynucleotide sequences encoding mammalian
 CC pheromone receptor polypeptides. The polypeptides are expressed in
 CC murine and rat vomeronasal organ. The products can be used for modifying
 CC pheromone activity, e.g. for decreasing pheromone receptor mediated
 CC signal transduction. They can be used for controlling fertility and
 CC behaviour in vertebrates and invertebrates. Compositions comprising the
 CC polypeptides are particularly useful in e.g. controlling fertility in
 CC livestock and controlling reproduction in rodents or insects by

CC interrupting the normal behaviours of rodents or insects that result in
CC reproduction. The present sequence represents the coding sequence of
CC mouse pheromone receptor VR10.

XX Sequence 1521 BP; 489 A; 264 C; 288 G; 480 T; 0 other;
XX Query Match 0.5%; Score 19; DB 20; Length 1521;
XX Best Local Similarity 100.0%; Pred. No. 1e+02;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2806 TGATATGCCCATGAGTGAT 2824
Db 1152 TGATATGCCCATGAGTGAT 1170
|||||

RESULT 38
ABL08786
ID ABL08786 standard; cDNA; 2606 BP.

XX ABL08786;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 20840.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WC200171042-A2.
XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.
XX P-PSDB; ABB64683.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1; SEQ ID NO 20840; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2606 BP; 760 A; 496 C; 506 G; 844 T; 0 other;

XX Query Match 0.5%; Score 19; DB 23; Length 2606;
XX Best Local Similarity 100.0%; Pred. No. 1e+02;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2262 CTAGATCAAGATATTAT 2280
Db 338 CTAGATCAAGATATTAT 356
|||||

RESULT 39
AA05810
ID AA05810 standard; cDNA; 2732 BP.
XX AA05810;
XX 04-MAY-1999 (first entry)
XX Mouse pheromone receptor VR10 encoding cDNA.
XX Pheromone receptor; signal transduction; fertility; behaviour;
KW reproduction; rodent; insect; mouse; ss.

XX Mus sp.
XX Key Location/Qualifiers
XX CDS 80..1378
XX FT /*tag= a

XX WO9900422-A1.
XX 07-JAN-1999.
XX 30-JUN-1998; 98WO-US13680.
XX 30-JUN-1997; 97US-0051284.
XX (HARD) HARVARD COLLEGE.

XX Buck L, Dulac C, Herrada G, Matsunami H;
XX WPI; 1999-095684/08.
XX P-PSDB; AAW94910.

XX New isolated pheromone receptor polypeptides - used to develop
PT products for controlling fertility and behaviour in vertebrates and
PT invertebrates

XX Claim 18; Page 99-101; 308pp; English.

XX The invention relates to polynucleotide sequences encoding mammalian
CC pheromone receptor polypeptides. The polypeptides are expressed in
CC murine and rat vomeronasal organ. The products can be used for modifying
CC pheromone activity, e.g. for decreasing pheromone receptor mediated
CC signal transduction. They can be used for controlling fertility and
CC behaviour in vertebrates and invertebrates. Compositions comprising the
CC polypeptides are particularly useful in e.g. controlling fertility in
CC livestock and controlling reproduction in rodents or insects by
CC interrupting the normal behaviours of rodents or insects that result in
CC reproduction. The present sequence represents a cDNA encoding a mouse
CC pheromone receptor VR10. The cDNA is deposited under the Genbank
CC accession number AF011420.

XX Sequence 2732 BP; 818 A; 532 C; 509 G; 873 T; 0 other;

XX Query Match 0.5%; Score 19; DB 20; Length 2732;

XX Best Local Similarity 100.0%; Pred. No. 1e+02;

XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2806 TGATATGCCCATGAGTGAT 2824
Db 1231 TGATATGCCCATGAGTGAT 1249
|||||

RESULT 40
AA05811
ID AA05811 standard; cDNA; 2962 BP.

XX AA05811;
XX 04-MAY-1999 (first entry)

```

XX Mouse pheromone receptor VR11 encoding cDNA.
DE
XX
XX Pheromone receptor; signal transduction; fertility; behaviour;
KW reproduction; rodent; insect; mouse; ss.
XX
XX Mus sp.
OS
XX
XX Key Location/Qualifiers
FH 81..1604
FT CDS /*tag= a
FT
XX WO9900422-A1.
XX
XX 07-JAN-1999.
XX
XX 30-JUN-1998; 98WO-US13680.
XX
XX 30-JUN-1997; 97US-0051284.
XX (HARD ) HARVARD COLLEGE.
XX
XX Buck L, Dulac C, Herrada G, Matsunami H;
PI WPI; 1999-095684/08.
DR P-PSDB; AAW94911.
XX
XX New isolated pheromone receptor polypeptides - used to develop
PT products for controlling fertility and behaviour in vertebrates and
PT invertebrates
XX
XX Claim 18; Page 102-105; 308pp; English.
XX
XX The invention relates to polynucleotide sequences encoding mammalian
CC pheromone receptor polypeptides. The polypeptides are expressed in
CC murine and rat vomeronasal organ. The products can be used for modifying
CC pheromone activity, e.g. for decreasing pheromone receptor mediated
CC signal transduction. They can be used for controlling fertility and
CC behaviour in vertebrates and invertebrates. Compositions comprising the
CC polypeptides are particularly useful in e.g. controlling fertility in
CC livestock and controlling reproduction in rodents or insects by
CC interrupting the normal behaviours of rodents or insects that result in
CC reproduction. The present sequence represents a cDNA encoding a mouse
CC pheromone receptor VR11. The cDNA is deposited under the Genbank
CC accession number AF011421.
XX
XX Sequence 2962 BP; 898 A; 567 C; 559 G; 938 T; 0 other;
SQ
Query Match 0.5%; Score 19; DB 20; Length 2962;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2806 TGATATGCCCATGAGTGAT 2824
DB 1232 TGATATGCCCATGAGTGAT 1250
|||||
RESULT 41
AAV36034/c
ID AAV36034 standard; DNA; 4446 BP.
XX
XX AAV36034;
AC
XX 02-SEP-1998 (first entry)
DT
XX DNA encoding a DNA polymerase of Pyrococcus sp. GE23.
DE
XX Heat stable DNA polymerase; archaeobacteria; secondary structure; ds.
KW
XX Pyrococcus sp.
OS
XX Key Location/Qualifiers
FH 1547..3862
FT CDS

```

```

FT
XX
XX WO9801567-A2.
XX
XX 15-JAN-1998.
XX
XX 10-JUL-1997; 97WO-FR01259.
XX
XX 10-JUL-1996; 96FR-0008631.
XX (APPL-) APPLIGENE-ONCOR SA.
XX
XX Cambon MA, Querellou J;
PI WPI; 1998-101062/09.
DR P-PSDB; AAW60719.
XX
XX Heat stable DNA polymerase from Pyrococcus species - and related
PT DNA, vectors and transformed cells, useful in nucleic acid
PT amplification reactions performed at temperatures high enough to
PT denature secondary structures
XX
XX Claim 9; Pages 37-41; 80pp; French.
XX
XX The present sequence encodes a heat stable DNA polymerase from
CC archaeobacteria of the genus Pyrococcus sp. GE23. The encoded polymerase
CC has an extremely high heat stability and can catalyse reactions at
CC 70-90 degrees Celsius. High molecular weight DNA was isolated from the
CC bacteria, and screened with probes based on the polymerase genes of
CC P. furiosus and Thermococcus litoralis to isolate the present sequence.
CC Host cells transformed with DNA encoding the present polymerase are
CC used to produce recombinant DNA which is useful for nucleic acid
CC amplification, including direct, reversed polymerase chain reaction (PCR)
CC on double-stranded matrices and PCR on matrices having secondary
CC structures that block the process at conventional elongation
CC temperatures.
XX
XX Sequence 4446 BP; 1358 A; 884 C; 1115 G; 1089 T; 0 other;
SQ
Query Match 0.5%; Score 19; DB 19; Length 4446;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 221 TTCCCTTATCTCTCTCTG 239
DB 1852 TTCCCTTATCTCTCTCTG 1834
|||||
RESULT 42
AAL36325
ID AAL36325 standard; DNA; 4656 BP.
XX
XX AAL36325;
AC
XX 08-JAN-2002 (first entry)
DT
XX Human musculoskeletal system related polynucleotide SEQ ID NO 2690.
DE
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
XX Homo sapiens.
OS
XX WO200155367-A1.
PN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US01338.
PF

```

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229387.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232599.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234397.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451937/48.
XX Isolated polypeptide for treating, preventing and/ or prognosing
XX disorders related to the musculoskeletal system including
PT

PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX Example 2: SEQ ID NO 2690; 781pp + Sequence Listing; English.
 PS
 XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 4656 BP; 1461 A; 810 C; 804 G; 1581 T; 0 other;
 Query Match 0.5%; Score 19; DB 22; Length 4656;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3567 TATATAATGTAATAGGGTG 3585
 Db 2822 TATATAATGTAATAGGGTG 2840
 |||||
 RESULT 43
 AAD11120/c
 ID AAD11120 standard; DNA; 4877 BP.
 AC AAD111120;
 XX
 DT 24-SEP-2001 (first entry)
 DE Human small cell lung cancer associated gene, KIAA0963.
 XX
 KW Human; small cell lung cancer; therapy; hCAAP; nucleic acid; NA;
 KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
 KW leiomyosarcoma; synovial sarcoma; cytostatic; ds.
 XX
 OS Homo sapiens.
 XX
 WO200153349-A2.
 PN
 PD 26-JUL-2001.
 XX
 PF 19-JAN-2001; 2001WO-US02015.
 XX
 PR 21-JAN-2000; 2000US-0489101.
 XX
 PA (LUDWIG) LUDWIG INST CANCER RES.
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
 XX
 WIPI; 2001-457597/49.
 XX
 DR Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAAP e.g. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 XX
 PS Claim 57: Page 105-107; 152pp; English.

XX The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterised by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterised by
 CC expression of a hCAAP, and determine regression, progression or onset
 CC of a condition characterised by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is a small cell lung cancer associated gene
 CC designated as NY-SCLC-11, encoding human KIAA0963 protein.
 XX
 SQ Sequence 4877 BP; 867 A; 1682 C; 1570 G; 758 T; 0 other;
 Query Match 0.5%; Score 19; DB 22; Length 4877;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2775 ATGCCCATGTCGATTGTC 2793
 Db 3238 ATGCCCATGTCGATTGTC 3220
 |||||
 RESULT 44
 AAC75998/c
 ID AAC75998 standard; cDNA; 4924 BP.
 XX AAC75998;
 AC AAC75998;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF1553 polynucleotide sequence SEQ ID NO:3105.
 XX
 KW Human; open reading frame; OREF; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nocrotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 WO200058473-A2.
 PN
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 WIPI; 2000-602362/57.
 DR P-PSDB; AAB41789.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame x,
 PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 5; Page 2323-2326; 5507pp; English.
 XX AACT74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antididiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX Sequence 4924 BP; 892 A; 1694 C; 1572 G; 765 T; 1 other;
 XX
 Query Match 0.5%; Score 19; DB 21; Length 4924;
 Best Local Similarity 100.0%; Pred. No. 1e-02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2775 ATGCCCATGTCGTATTTC 2793
 Db 3239 ATGCCCATGTCGTATTTC 3221
 RESULT 45
 ABL30262/c
 ID ABL30262 standard; DNA; 5415 BP.
 XX
 AC ABL30262;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 42259.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 KW
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR
 XX 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Claim 1; SEQ ID NO 42259; 21pp + Sequence Listing; English.
 PS
 XX
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 5415 BP; 1512 A; 1457 C; 1267 G; 1179 T; 0 other;
 XX
 Query Match 0.5%; Score 19; DB 23; Length 5415;
 Best Local Similarity 100.0%; Pred. No. 1.1e-02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3562 ATATATATATAATGTAATA 3580
 Db 5286 ATATATATATAATGTAATA 5268
 RESULT 46
 AAH62719
 ID AAH62719 standard; DNA; 18234 BP.
 XX
 AC AAH62719;
 XX
 DT 11-SEP-2001 (first entry)
 XX
 DE Shrimp white spot Bacilliform virus (WSBV) gene 30.
 XX
 XX Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
 KW antiviral agent; gene expression; antisense construct;
 KW transgenic viral resistant shrimp; ds.
 KW
 XX White spot syndrome virus.
 OS
 XX WO200138351-A2.
 PN
 XX 31-MAY-2001.
 PD
 XX
 XX 08-NOV-2000; 2000WO-US28888.
 PF
 XX 24-NOV-1999; 99CN-0124717.
 XX
 XX (PENY-) PE CORP NY.
 PA
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
 PA (SINO-) SINOGENOMAX CO LTD.
 XX
 XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
 PI
 XX WPI; 2001-355877/37.
 DR
 XX P-PSDB; AAG84939.
 DR
 XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
 PT (WSBV), useful for producing viral polypeptides that can be used to
 PT screen for agents that are useful for treating WSBV infection -
 XX
 XX Claim 4; Figure 3; 626pp; English.
 PS
 XX The invention provides the primary nucleotide sequence of the WSBV genome
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
 CC encoded proteins (AAG84910-AAG85031) and oligonucleotide sequences
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
 CC molecules and proteins of the invention are useful for diagnosis and
 CC monitoring viral infection, in screens for antiviral agents and for
 CC monitoring viral gene expression or activity during a treatment regimen.
 CC The nucleic acid molecules are also useful as antisense constructs to
 CC control viral gene expression in infected cells and tissues and to create
 CC transgenic viral resistant shrimp.
 XX

```

SQ Sequence 18234 BP; 5373 A; 4270 C; 3810 G; 4781 T; 0 other;
Query Match 0.5%; Score 19; DB 22; Length 18234;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 ATTCAATTGGCATTGGAT 572
Db 4662 ATTCAATTGGCATTGGAT 4680

RESULT 47
AAH62689
ID AAH62689 standard; DNA; 305107 BP.
XX
AC AAH62689;
XX
DT 11-SEP-2001 (first entry)
XX
DE Shrimp white spot Bacilliform virus (WSBV) genomic sequence.
XX
KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW antiviral agent; gene expression; antisense construct;
KW transgenic viral resistant shrimp; ds.
XX
OS White spot syndrome virus.
XX
PN WO200138351-A2.
XX
PD 31-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US28888.
XX
PR 24-NOV-1999; 95CN-0124717.
XX
PA (PENY-) PE CORP NY.
XX
PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX
PI (SINO-) SINOGENOMAX CO LTD.
XX
PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX WPI; 2001-355877/37.
XX
PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT (WSBV), useful for producing viral polypeptides that can be used to
PT screen for agents that are useful for treating WSBV infection.
XX
PS Disclosure; Figure 2; 626pp; English.
XX
CC The invention provides the primary nucleotide sequence of the WSBV genome
CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC molecules and proteins of the invention are useful for diagnosis and
CC monitoring viral infection, in screens for antiviral agents and for
CC monitoring viral gene expression or activity during a treatment regimen.
CC The nucleic acid molecules are also useful as antisense constructs to
CC control viral gene expression in infected cells and tissues and to create
CC transgenic viral resistant shrimp.
XX
SQ Sequence 305107 BP; 92042 A; 62482 C; 62635 G; 87948 T; 0 other;
Query Match 0.5%; Score 19; DB 22; Length 305107;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 ATTCAATTGGCATTGGAT 572
Db 214277 ATTCAATTGGCATTGGAT 214295

RESULT 48
AAD04376/c

```

```

ID AAD04376 standard; DNA; 35 BP.
XX
AC AAD04376;
XX
DT 04-JUL-2001 (first entry)
XX
DE Reverse PCR primer, P127BGL to amplify tomato RdRP cDNA.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
KW PCR primer; ss.
XX
OS Lycopersicon esculentum.
XX
PN US6218142-B1.
XX
PD 17-APR-2001.
XX
PF 05-MAR-1997; 97US-0811583.
XX
PR 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASENEGGER M.
XX
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX WPI; 2001-289830/30.
XX
PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding.
XX
PS Example 3; Column 21; 31pp; English.
XX
CC The present sequence is reverse PCR primer, P127BGL, used to amplify a
CC cDNA encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)
CC activity. This primer contains BglII site to enhance the cloning
CC efficiency. This protein is capable of RNA-directed RNA synthesis, thus
CC using RNA as a template for synthesizing complementary RNA molecules.
CC RdRP nucleic acid is useful for modulating gene expression in plants,
CC humans and animals. This may lead to various physiological, developmental
CC and/or morphological changes. Transgenic plants containing RdRP nucleic
CC acid is especially useful in plant cell or tissue cultures and in plant
CC breeding. RdRP is useful in gene therapy, particularly for treating a
CC disease that is caused by the undesirable expression or overexpression of
CC a gene.
XX
SQ Sequence 35 BP; 13 A; 7 C; 8 G; 7 T; 0 other;
Query Match 0.5%; Score 18; DB 22; Length 35;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3615 GACTACCTTTGCTTTA 3632
Db 35 GACTACCTTTGCTTTA 18

RESULT 49
AAS17843/c
ID AAS17843 standard; DNA; 35 BP.
XX
AC AAS17843;
XX
DT 08-MAY-2002 (first entry)
XX
DE Tomato RNA-directed RNA polymerase (RdRP) P127Bgl PCR primer.
XX
KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytosstatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection; RACE;

```

KW immunoprecipitation; immunolocalisation; PCR primer; p127Bgl;
 KW gene therapy; ss.

XX Lycopersicon esculentum.

XX US2001023067-A1.

XX 20-SEP-2001.

XX 08-FEB-2001; 2001US-0782874.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASSENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-595798/67.

XX New nucleic acid molecule encoding a polypeptide having the enzymatic
 PT activity of RNA-directed RNA polymerase, for modulating gene expression
 PT and treating cancer and virus infection in human and animals

XX Example 3; Page 12; 34pp; English.

XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
 CC specific PCR primer p127Bgl used to clone the full length RdRP cDNA
 CC molecule of the invention. The invention comprises the nucleic acid and
 CC protein sequences of RdRP, the protein of the invention can catalyse in
 CC vitro transcription of short single stranded RNAs into DNA molecules,
 CC this transcription can be either primed by RNA or DNA oligonucleotides,
 CC or be unprimed. The protein may have cytosolic or virucide activities.
 CC The sequences of the invention may be used in gene therapy or as an RNA
 CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template
 CC nucleic acid molecule derived from a nucleic acid molecule which causes
 CC a disease are useful for treating a disease caused by the undesired
 CC expression or overexpression of a nucleic acid molecule in a human, rat
 CC or mouse, by administering the molecules. This system can be used in the
 CC preparation of a pharmaceutical composition and for inhibiting
 CC expression of any desired gene by transferring the RdRP system to
 CC organisms that either lack a comparable mechanism or do not sufficiently
 CC express their own RdRP. An antibody or an antagonist or inhibitor to the
 CC protein are useful for inhibiting RNA directed RNA synthesis and for
 CC ensuring stable heterologous, gene expression in transgenic organisms.
 CC The sequence is useful for probes and/or for the control of gene
 CC expression, as primers for amplification of nucleic acid molecules and
 CC as tools for the detection of expression of the cDNA molecules.
 CC Additionally, nucleotide and protein sequences are useful for
 CC suppression of undesired gene expression in humans and animals. The RdRP
 CC is useful as a therapeutic agent for the control of cancer and virus
 CC infection in humans and animals and the antibody is useful for
 CC immunoprecipitation or immunolocalisation of the protein, identification
 CC of polypeptides interacting with it and screening expression libraries.

XX Sequence 35 BP; 13 A; 7 C; 8 G; 7 T; 0 other;

Query Match 0.5%; Score 18; DB 23; Length 35;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3615 GACTACCTTTGTCTTTA 3632

DB 35 GACTACCTTTGTCTTTA 18
 |||||

RESULT 50

RAF29312

ID RAF29312 standard; DNA; 48 BP.

XX AC

XX AAF29312;

XX DT

18-APR-2001 (first entry)

XX

DE Primer base sequence used to illustrate primer selection method.

XX

KW Primer; optimum sequence; differential display; ss.

XX

OS Synthetic.

XX

PN JP2000308487-A.

XX

PD 07-NOV-2000.

XX

XX 30-MAR-1999; 95JP-0088410.

XX

XX 30-MAR-1999; 95JP-0088410.

XX

PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX

WPI; 2001-046077/06.

XX

PT Selection of primer base for optimizing primer selection comprises
 PT obtaining an optimum sequence for differential display from an
 PT expression gene data base

XX

PS Disclosure; Fig 9; 13pp; Japanese.

XX

CC This invention relates to a method for selecting the sequence of a
 CC primer. The method comprises obtaining an optimum sequence for
 CC differential display from an expression gene data base, and using the
 CC base sequences most frequently expressed as the primer candidates in the
 CC order of frequency. The optimum primer group characterised by the use of
 CC genetic algorithm from the primer candidates is selected. The method is
 CC used for selecting a primer sequence quickly. The present sequence
 CC represents a primer used in an illustration of the method of the
 CC invention.

XX

SQ Sequence 48 BP; 18 A; 2 C; 5 G; 23 T; 0 other;

Query Match 0.5%; Score 18; DB 22; Length 48;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 571 ATTGAAGAAATTTTTTTT 588

DB 5 ATTGAAGAAATTTTTTTT 22
 |||||

Search completed: November 5, 2002, 23:52:42
 Job time : 1690 secs

FEATURE:
NAME/KEY: CDS
LOCATION: 194...3535
US-08-811-583-1

Query Match 100.0%; Score 3731; DB 4; Length 3731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATATTCCTTACTTACTTACCAGGATGACTCATCACTCCCTCAAGTCTTTGTGT 60
DB 1 GAAATATTCCTTACTTACTTACCAGGATGACTCATCACTCCCTCAAGTCTTTGTGT 60

QY 61 GTTGTGAATAAATTTGGTTGGCTTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT 120
DB 61 GTTGTGAATAAATTTGGTTGGCTTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT 120

QY 121 TTTGCAFAACTTCAGGGGATTCAGTTGGTTTGGTTTGGATTGAAAGTGGAACTGCACAT 180
DB 121 TTTGCAFAACTTCAGGGGATTCAGTTGGTTTGGTTTGGATTGAAAGTGGAACTGCACAT 180

QY 181 GGAATTTGGCTACATGGGAAAGACAATTCAGGTTTTCGGATTCCTTCCTCTCTGC 240
DB 181 GGAATTTGGCTACATGGGAAAGACAATTCAGGTTTTCGGATTCCTTCCTCTCTGC 240

QY 241 GGAAGTGGTTAAGTCATCTTACAGAAATATACAGATATGGAACGTATGTCATTGGA 300
DB 241 GGAAGTGGTTAAGTCATCTTACAGAAATATACAGATATGGAACGTATGTCATTGGA 300

QY 301 GGTAAACAGTCCAAAGGAGATCTAGACATTTGCCAAGTTCAATTTGCCGACAAACAT 360
DB 301 GGTAAACAGTCCAAAGGAGATCTAGACATTTGCCAAGTTCAATTTGCCGACAAACAT 360

QY 361 AAGTGTGCAGAAATCATCTACTTGGCTATACAGGCTGTATTTGGCTCTCTCTATTT 420
DB 361 AAGTGTGCAGAAATCATCTACTTGGCTATACAGGCTGTATTTGGCTCTCTCTATTT 420

QY 421 GAAGCTTGGGAAATGAAACTGATATGTCGAACCTCGGGCATATGTCGATCAGATGGA 480
DB 421 GAAGCTTGGGAAATGAAACTGATATGTCGAACCTCGGGCATATGTCGATCAGATGGA 480

QY 481 TGGCATTAATTTGAAATTCGGATGTCAGATATCAGATGACAACTTTGCAAGTGTGGGAG 540
DB 481 TGGCATTAATTTGAAATTCGGATGTCAGATATCAGATGACAACTTTGCAAGTGTGGGAG 540

QY 541 TACAGAAGTTTCAATTCGAATTTGGCATTTGAAAGAAATTTTCTTTTATCTAG 600
DB 541 TACAGAAGTTTCAATTCGAATTTGGCATTTGAAAGAAATTTTCTTTTATCTAG 600

QY 601 TGGTTCAGCTGACTATAAATTCAGCTTTCATATGAAATATATGCGAGTTTGTCTCCA 660
DB 601 TGGTTCAGCTGACTATAAATTCAGCTTTCATATGAAATATATGCGAGTTTGTCTCCA 660

QY 661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCTCATACAGTTATTTGGTCTCTCTCGAT 720
DB 661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCTCATACAGTTATTTGGTCTCTCTCGAT 720

QY 721 CTATAGAGACTTGAAGAACTCCCTGTTATAGCTTCTTTAAGGAACTCCTGATCAGTG 780
DB 721 CTATAGAGACTTGAAGAACTCCCTGTTATAGCTTCTTTAAGGAACTCCTGATCAGTG 780

QY 781 GGTGAGGACACAGATTTCCCTCCATCTGATAGGCTATCTTCTAGCTTTATGTTGCA 840
DB 781 GGTGAGGACACAGATTTCCCTCCATCTGATAGGCTATCTTCTAGCTTTATGTTGCA 840

QY 841 GTTCCGTAGGGGTGCTCTTCCAAATTTTCAGGAAAGTTTTTCCACTATGCAAGC 900
DB 841 GTTCCGTAGGGGTGCTCTTCCAAATTTTCAGGAAAGTTTTTCCACTATGCAAGC 900

QY 901 TGAAACAATATTAATTTACAGCTGTTTCACTTTTGGTCTCTCAAAAATTCGGCTCT 960
DB 901 TGAAACAATATTAATTTACAGCTGTTTCACTTTTGGTCTCTCAAAAATTCGGCTCT 960

QY 961 GGTTCCTCCCAATGTCCAGCCTCCGGAAGAAATTTCAATTCCTTACAGATTTTGTTCAAAAT 1020
DB 961 GGTTCCTCCCAATGTCCAGCCTCCGGAAGAAATTTCAATTCCTTACAGATTTTGTTCAAAAT 1020

QY 1021 TAGTTCTTTTGTACAGCATGGATGATACCTGGCCAGCATTAATGATGCTACTTTTTCCG 1080
DB 1021 TAGTTCTTTTGTACAGCATGGATGATACCTGGCCAGCATTAATGATGCTACTTTTTCCG 1080

QY 1081 ATTAGTTGATCCTCGAAGGAGAAATGTGGCATGATGAGCATGCGCTTAGAGAACTGTA 1140
DB 1081 ATTAGTTGATCCTCGAAGGAGAAATGTGGCATGATGAGCATGCGCTTAGAGAACTGTA 1140

QY 1141 CTATATAAAGGAGTCTGTTATGATCCCGTGGGTGGCTCACTGAGCAGTATGATGGTA 1200
DB 1141 CTATATAAAGGAGTCTGTTATGATCCCGTGGGTGGCTCACTGAGCAGTATGATGGTA 1200

QY 1201 TCTCAAGGATAGACAACCTCCAAAATCTCCGTCCATCACTTTAGATGATGGTTGGTGA 1260
DB 1201 TCTCAAGGATAGACAACCTCCAAAATCTCCGTCCATCACTTTAGATGATGGTTGGTGA 1260

QY 1261 TGTAGAAGGCTCTAGTAACACCATGCAAGTTTATTTTGTGTCCAGAGGTTAATGT 1320
DB 1261 TGTAGAAGGCTCTAGTAACACCATGCAAGTTTATTTTGTGTCCAGAGGTTAATGT 1320

QY 1321 TTCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAAATTTCTTCTGTTCTTT 1380
DB 1321 TTCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAAATTTCTTCTGTTCTTT 1380

QY 1381 TGTGATGAGGAGTGGGAGAACTGTATCTACAGACTTATACCAAGCAAGTACTGG 1440
DB 1381 TGTGATGAGGAGTGGGAGAACTGTATCTACAGACTTATACCAAGCAAGTACTGG 1440

QY 1441 AAGTGGTGTACGACAAAACATCTATGAGAGATCTTATCAACTCTCGGAAAGGCTTTGT 1500
DB 1441 AAGTGGTGTACGACAAAACATCTATGAGAGATCTTATCAACTCTCGGAAAGGCTTTGT 1500

QY 1501 AATTTGGTGAATAAATTTGAAATTTCTTGCATTTTCATCGAGCCAGTTGCGGGATATTC 1560
DB 1501 AATTTGGTGAATAAATTTGAAATTTCTTGCATTTTCATCGAGCCAGTTGCGGGATATTC 1560

QY 1561 AGTGTGGATTTTGGCATCAAGACCTGGCTTACTCAAAATGATATAAGAGCTTTGATGGG 1620
DB 1561 AGTGTGGATTTTGGCATCAAGACCTGGCTTACTCAAAATGATATAAGAGCTTTGATGGG 1620

QY 1621 TCATTTTTCGAGATCAAGATGTCGAAAATATGCTGCCAGACTTGGTCAATCTTTTGG 1680
DB 1621 TCATTTTTCGAGATCAAGATGTCGAAAATATGCTGCCAGACTTGGTCAATCTTTTGG 1680

QY 1681 TTCCCTCCAGAGAGCTTTGAGTGTCTTAGGCATGAGATTGAAGTTATCCCGATGAAA 1740
DB 1681 TTCCCTCCAGAGAGCTTTGAGTGTCTTAGGCATGAGATTGAAGTTATCCCGATGAAA 1740

QY 1741 GGTTCATGGAACCCAGCTATGCTTTTCTGATGGAATTTGTAATAATCTGAGTCTTTGC 1800
DB 1741 GGTTCATGGAACCCAGCTATGCTTTTCTGATGGAATTTGTAATAATCTGAGTCTTTGC 1800

QY 1801 TCATAGAGTTGGCTCAAAATGTCGCTTCAATATACCCATCTGCTTCCAGATTCGTTA 1860
DB 1801 TCATAGAGTTGGCTCAAAATGTCGCTTCAATATACCCATCTGCTTCCAGATTCGTTA 1860

QY 1861 TGGTGGATATAAAGTGTGTGGGTGTTGATCCGGATTCATCAATGAAGTTTCTCTTTGAG 1920
DB 1861 TGGTGGATATAAAGTGTGTGGGTGTTGATCCGGATTCATCAATGAAGTTTCTCTTTGAG 1920

QY 1921 AAAGAGCATGCGAAATATGAATCAGACAACTAAAGTAGATGCTTGGATGGAGAA 1980
DB 1921 AAAGAGCATGCGAAATATGAATCAGACAACTAAAGTAGATGCTTGGATGGAGAA 1980

QY 1981 ATATCAGCTTGTATCTTAAATCGTCACTGATTAACGTTCTTGTCTACACTTTGAGTGA 2040
DB 1981 ATATCAGCTTGTATCTTAAATCGTCACTGATTAACGTTCTTGTCTACACTTTGAGTGA 2040

QY 2041 AGATGAAGTTCTCGAACAGAAAGCAAAAGGAACTGTAGATCAGCTTGATCTATCTTGA 2100

Db	2041	AGATGAAGTTCTGACAGAGACAAAGAGAGCTTAGATCAGCTTGATGCTATCTTTGCA	2100
QY	2101	TGATTCCTTTAAAGGCACAGAGGCTTTGGAAATTGATGTCCTCTGGAGAGAACACTAATAT	2160
Db	2101	TGATTCCTTTAAAGGCACAGAGGCTTTGGAAATTGATGTCCTCTGGAGAGAACACTAATAT	2160
QY	2161	TCFCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTCAGGCCCTTCCTTCAATGAT	2220
Db	2161	TCFCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTCAGGCCCTTCCTTCAATGAT	2220
QY	2221	GTTGCAAACTCTCCGGCATCCAAAGTTGCTCGATTTGGGACTAGATCAAGAATATTTAT	2280
Db	2221	GTTGCAAACTCTCCGGCATCCAAAGTTGCTCGATTTGGGACTAGATCAAGAATATTTAT	2280
QY	2281	TCCAAATGGAAGAACAAATGATGGGATGTTTGGATCAATCCAGAACCTTTGGAATATGTCA	2340
Db	2281	TCCAAATGGAAGAACAAATGATGGGATGTTTGGATCAATCCAGAACCTTTGGAATATGTCA	2340
QY	2341	GGTGTGTTGTTCASTTTACTGGTGCTGGACATGGAGAGTTTCTGACGATTTACATCCATT	2400
Db	2341	GGTGTGTTGTTCASTTTACTGGTGCTGGACATGGAGAGTTTCTGACGATTTACATCCATT	2400
QY	2401	TAATAACAGCAGATCCACCACAGTAATTCATTCTGAAGGGAATGTGTGTTGCAAA	2460
Db	2401	TAATAACAGCAGATCCACCACAGTAATTCATTCTGAAGGGAATGTGTGTTGCAAA	2460
QY	2461	AAATCCATGCTTGCATCCTGGTGATATTGCTGTGTTTTAAAGGCTGTAAATGTCGAGGCT	2520
Db	2461	AAATCCATGCTTGCATCCTGGTGATATTGCTGTGTTTTAAAGGCTGTAAATGTCGAGGCT	2520
QY	2521	GCACCACATGGTAGATTGTGTTGATTCCTCTCAGAAAGAAAAGACCTCATCCGAATGA	2580
Db	2521	GCACCACATGGTAGATTGTGTTGATTCCTCTCAGAAAGAAAAGACCTCATCCGAATGA	2580
QY	2581	ATGTTCTGGAGTGATTGGATGGGGATATCTACTTTGCTTGGGATCAAGACATGAT	2640
Db	2581	ATGTTCTGGAGTGATTGGATGGGGATATCTACTTTGCTTGGGATCAAGACATGAT	2640
QY	2641	CCGCGCAAGCGAAGTCCAGCCGATGGAATATCCTCCAGACCCAGCATACAGTTGCAACA	2700
Db	2641	CCGCGCAAGCGAAGTCCAGCCGATGGAATATCCTCCAGACCCAGCATACAGTTGCAACA	2700
QY	2701	TGATGTCACAAATTGAGGAAGTTGAGAGTACTTCCACCACTATATTGTGGATGACAGTTT	2760
Db	2701	TGATGTCACAAATTGAGGAAGTTGAGAGTACTTCCACCACTATATTGTGGATGACAGTTT	2760
QY	2761	GGGAATCATAGCAAAATGCCATGTCGTATTTGCAGACAGAGAACTGATGGCCATGAG	2820
Db	2761	GGGAATCATAGCAAAATGCCATGTCGTATTTGCAGACAGAGAACTGATGGCCATGAG	2820
QY	2821	TGATCATCGAAAAAATGCTGAGCTCTTTTCAATTGCAATGGACITTTCCAAAGACTGG	2880
Db	2821	TGATCATCGAAAAAATGCTGAGCTCTTTTCAATTGCAATGGACITTTCCAAAGACTGG	2880
QY	2881	TGTTCCGCGTGAATACCATCTCAGTTGCGGCCCTTAAGAAATACCAGACTTCATGGATAA	2940
Db	2881	TGTTCCGCGTGAATACCATCTCAGTTGCGGCCCTTAAGAAATACCAGACTTCATGGATAA	2940
QY	2941	GCGGACACAGACCGGTATATCTCAGAAGAGTATTATGGAAAGCTTTTCAGGAAAGTGAA	3000
Db	2941	GCGGACACAGACCGGTATATCTCAGAAGAGTATTATGGAAAGCTTTTCAGGAAAGTGAA	3000
QY	3001	GGACAAGCACCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGTTGCAAGGAGATC	3060
Db	3001	GGACAAGCACCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGTTGCAAGGAGATC	3060
QY	3061	ATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTTACATTGACCAAGCTTTTGACTA	3120
Db	3061	ATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTTACATTGACCAAGCTTTTGACTA	3120
QY	3121	CAAAACCTGAATATGACAACAAGCTGGGTAAATTTAATGAGCTACTATTGGCATAAAAACAGA	3180

```
; REFERENCE/DOCKET NUMBER: 7555-00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (313) 641-1600
; TELEFAX: (313) 641-0270
; TELEX: 287637 Harness UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Solanum tuberosum
; INDIVIDUAL ISOLATE: p-PCM-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 81..530
US-08-100-874-1

Query Match          0.8%; Score 30; DB 1; Length 906;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3687 TAAGTACTATATGATGTGATGTTGAGTTT 3716
      ||||||||||||||||||||||||||||
Db 47 TAAGTACTATATGATGTGATGTTGAGTTT 18

RESULT 3
US-08-811-583-4/c
; Sequence 4, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
US-08-811-583-6

Query Match          0.7%; Score 25; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GCCGATGGATATCTCCAGCACCC 2683
      ||||||||||||||||||||||||
Db 25 GCCGATGGATATCTCCAGCACCC 1

RESULT 5
US-08-811-583-5/c
; Sequence 5, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
```


NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 1750 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HMC1NOT01
CLONE: 9476
US-09-234-613-54

Query Match 0.5%; Score 20; DB 3; Length 1750;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGCGAA 1935
|||||
DB 1355 TTGAGAAAGAGCATGCGAA 1374

RESULT 8

US-08-811-583-7
Sequence 7, Application US/08811583
Patent No. 6218142

GENERAL INFORMATION:

APPLICANT: Wassenecker, Michael
APPLICANT: Riedel, Leonhard
APPLICANT: Schiebel, Winfried
APPLICANT: Sanger, Heinz
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
RNA-DIRECTED RNA POLYMERASE (RdRP)
TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
RNA-DIRECTED RNA POLYMERASE (RdRP)
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:

NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 34 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "primer"

US-08-811-583-7

Query Match 0.5%; Score 19; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ACTCATCACTCCCTCAAG 51
|||||
DB 16 ACTCATCACTCCCTCAAG 34

RESULT 9

US-08-990-823-5/c
Sequence 5, Application US/08990823D
Patent No. 6228371

GENERAL INFORMATION:

APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
Immunostimulatory Peptides
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 49086
CURRENT APPLICATION NUMBER: US/08/990,823D
CURRENT FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: US 96/10375
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000,254
EARLIER FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 456
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis

FEATURE:

NAME/KEY: Modified base

OTHER INFORMATION: n represents a or g or c or t/u

US-08-990-823-5

Query Match 0.5%; Score 19; DB 4; Length 456;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2406 ACAGCAGATCCCAACAG 2424
|||||
DB 296 ACAGCAGATCCCAACAG 278

RESULT 10

US-07-732-242C-8/c
Sequence 8, Application US/07732242C
Patent No. 5298399

GENERAL INFORMATION:

APPLICANT: Uozumi, Takeshi; Masaki, Haruhiko;
APPLICANT: Hidaka, Makoto; Nakamura, Akira;
APPLICANT: Maeda, Michihisa; Yoneta, Yasuo
TITLE OF INVENTION: Gene of Urease
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Frisnauf, Holtz, Goodman & Woodward, P.C.
STREET: 600 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10016-2088
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 0.72mb
COMPUTER: IBM PC compatible (NEC PC-9801 RX)
OPERATING SYSTEM: MS DOS
SOFTWARE: ASCII Form

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/732,242C
FILING DATE: 19910718
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JPN 2-210178
FILING DATE: 10-AUG-1990

```

, ATTORNEY/AGENT INFORMATION:
,
, NAME: Goodman, Herbert
,
, REGISTRATION NUMBER: 17081
,
, REFERENCES/DOCKET NUMBER: 910532/HG
,
, TELECOMMUNICATION INFORMATION:
,
, TELEPHONE: (212)972-1400
,
, TELEFAX: (212)370-1622
,
, TELEX: 236268
,
, INFORMATION FOR SEQ ID NO: 8:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 6131 base pairs
,
, TYPE: NUCLEIC ACID
,
, STRANDEDNESS: both
,
, TOPOLOGY: linear
,
, MOLECULE TYPE: DNA (genomic)
,
US-07-732-242C-8

```

```
Query Match          0.5%; Score 19; DB 1; Length 6131;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19: Conservative 0; Mismatches 0; Indels 0; Caps 0;
```

Qy	922	GACTGGTTTCACCTTTTC	940
Db	2524	GACTGGTTTCACCTTTTC	2506

RESULT 11
US-09-813-133A-3/c
; Sequence 3, Application US/09813133A

```

1  GENERAL INFORMATION:
2  APPLICANT: GAN, weiniu et al
3  TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
4  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
5  TITLE OF INVENTION: USES THEREOF
6  FILE REFERENCE: CL001173
7  CURRENT APPLICATION NUMBER: US/09/813,133A
8  CURRENT FILING DATE: 2001-06-06
9  NUMBER OF SEQ ID NOS: 4
10 SOFTWARE: FastSeq for Windows Version 4.0
11 SEQ ID NO 3

```

```
Query Match          0.5%; Score 19; DB 4; Length 55827;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 627 TTTCATATGAAATATATG 645
|||||
Db 52586 TTTCATATGAAATATATG 52568

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/98/811,583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 356
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 35 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "primer"
US-08-a11-583-9

Query Match	0.5%	Score 18;	DB 4;	Length 35;
Best Local Similarity	100.0%;	Pred. No. 36;		
Matches 18:	Conservative	0;	Mismatches 0;	Indels 0;
	Gaps	0;		

Qy 3615 GACTACCTTTTGTCTTTA 3632
Db 35 GACTACCTTTTGTCTTTA 18

RESULT 13
US-09-058-389A-20/c
; Sequence 20, Application US/09058389A
; Patent No. 6130065
; GENERAL INFORMATION:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

ADDRESSES: David A. Jackson, Esq.
STREET: 411 Hackensack Ave., Continental Plaza, 4th
FLOOR
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/058,389A
FILING DATE: April 9, 1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1340-1-013N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800

```
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 461 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: other nucleic acid
;   DESCRIPTION: /desc = "intron 10(5')"
```

US-09-058-389A-20

Query Match 0.5%; Score 18; DB 3; Length 461;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2068 GGAAGCTGTAGTCAGCT 2085
|||||
DB 427 GGAAGCTGTAGTCAGCT 410

RESULT 14
US-09-611-781-20/C
; Sequence 20, Application US/09611781
; Patent No. 5423829
; GENERAL INFORMATION:
; APPLICANT: Belt, Judith A.
; APPLICANT: Crawford, Charles R.
; APPLICANT: Patel, Divyen
; TITLE OF INVENTION: A NITROBENZYL MERCAPTOPYRINERIBOSIDE
; TITLE OF INVENTION: (NEMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/611,781
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 09/058,389
; FILING DATE: April 9, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-013N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "intron 10(5')"

US-09-611-781-20

Query Match 0.5%; Score 18; DB 4; Length 461;

Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2068 GGAAGCTGTAGTCAGCT 2085
|||||
DB 427 GGAAGCTGTAGTCAGCT 410

RESULT 15
PCT-US95-08295-15
; Sequence 15, Application PC/TUS9508295
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: BREAST SPECIFIC GENES AND PROTEINS
; NUMBER OF SEQUENCES: 30
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08295
; FILING DATE: 30-JUN-1995
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US95-08295-15

Query Match 0.5%; Score 18; DB 5; Length 481;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2989 CAGGAAGCTGAAGGACAA 3006
|||||
DB 26 CAGGAAGCTGAAGGACAA 43

RESULT 16
US-09-641-638-345/C
; Sequence 345, Application US/09641638
; Patent No. 5432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR FILING DATE: 2000-02-11
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 345
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 501

; OTHER INFORMATION: 12-854-64 : polymorphic base A or G
; NAME/KEY: misc_binding
; LOCATION: 481..500
; OTHER INFORMATION: 12-854-64.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 502..521
; OTHER INFORMATION: 12-854-64.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 438..457
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 942..962
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 489..513
; OTHER INFORMATION: 12-854-64 potential probe
US-09-641-638-345

Query Match 0.5%; Score 18; DB 4; Length 1001;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3562 ATATATATATATGTAAT 3579

Db 182 ATATATATATGTAAT 165

RESULT 17

US-09-134-001C-1922
; Sequence 1922, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 1922

; LENGTH: 1440

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1922

Query Match 0.5%; Score 18; DB 4; Length 1440;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 661 TCGTCCATATGTCAGAAA 678

Db 603 TCGTCCATATGTCAGAAA 620

RESULT 18

US-08-383-750-3/C

; Sequence 3, Application US/08383750

; Patent No. 5744301

; GENERAL INFORMATION:

; APPLICANT: Birkenbach, Mark

; APPLICANT: Kieff, Elliott

; TITLE OF INVENTION: Epstein Barr Virus Induced Genes

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Avenue, N.W.,

; CITY: Suite 600

; STATE: Washington

; D.C.

; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,750
; FILING DATE: Herewith
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel, L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0827.3300001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1643 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..1116
US-08-383-750-3

Query Match 0.5%; Score 18; DB 1; Length 1643;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GTTGATGAGGAGTGGGAG 1399

Db 540 GTTGATGAGGAGTGGGAG 523

RESULT 19

US-08-352-678-3/C

; Sequence 3, Application US/08352678

; Patent No. 6043351

; GENERAL INFORMATION:

; APPLICANT: Birkenbach, Mark

; APPLICANT: Kieff, Elliott

; TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

; STREET: 600 Atlantic Avenue

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/352,678

; FILING DATE: 30-NOV-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/980,518

; FILING DATE: 25-NOV-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Gates, Edward R.

; REGISTRATION NUMBER: 31,616

; REFERENCE/DOCKET NUMBER: B0801/7044

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-720-3500

; TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1643 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 34...1116
US-08-352-678-3

Query Match 0.5%; Score 18; DB 3; Length 1643;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GTTGATGAGGAGTGGGAG 1399
|||||
DB 540 GTTGATGAGGAGTGGGAG 523

RESULT 20
PCT-US93-09636-3/c
Sequence 3, Application PC/TUS9309636
GENERAL INFORMATION:
APPLICANT: Birkenbach, Mark
TITLE OF INVENTION: Epstein Barr Virus Induced Genes
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09636
FILING DATE: herewith
CLASSIFICATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1643 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 34...1116
PCT-US93-09636-3

Query Match 0.5%; Score 18; DB 5; Length 1643;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GTTGATGAGGAGTGGGAG 1399
|||||
DB 540 GTTGATGAGGAGTGGGAG 523

RESULT 21
US-08-454-720A-35
Sequence 35, Application US/08454720A
Patent No. 5766888
GENERAL INFORMATION:

APPLICANT: Sobol, Robert E.
APPLICANT: Green, Mark R.
APPLICANT: Kawasaki, Ernest S.
TITLE OF INVENTION: Detection of Carcinoma Metastases by
TITLE OF INVENTION: Nucleic Acid Amplification
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffman-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.01, version
CURRENT APPLICATION DATA: US/08/454,720A
APPLICATION NUMBER: US/08/454,720A
FILING DATE: May 31, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/357,565
FILING DATE: December 16, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Petry, Douglas
REGISTRATION NUMBER: 35321
REFERENCE/DOCKET NUMBER: 9178
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2974
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 1990 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-454-720A-35
Query Match 0.5%; Score 18; DB 1; Length 1990;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 576 AGAAATTTTCTTTT 593
|||||
DB 1445 AGAAATTTTCTTTT 1462
RESULT 22
US-09-323-872A-45/c
Sequence 45, Application US/09323872A
Patent No. 6395539
GENERAL INFORMATION:
APPLICANT: Coschigano, Peter
TITLE OF INVENTION: Compositions and Methods for Bioremediation
FILE REFERENCE: OHU-03640
CURRENT APPLICATION NUMBER: US/09/323,872A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 09/072,433
PRIOR FILING DATE: 1998-05-04
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patent in version 3.0
SEQ ID NO 45
LENGTH: 2038
TYPE: DNA
ORGANISM: Thauera aromatica
US-09-323-872A-45
Query Match 0.5%; Score 18; DB 4; Length 2038;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1368 TTGCTGTTTCTTTTGTG 1385
|||||
Db 1212 TTGCTGTTTCTTTTGTG 1195

RESULT 23

US-09-620-412C-328/C
; Sequence 328, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 328
; LENGTH: 2148
; TYPE: DNA
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-328

Query Match 0.5%; Score 18; DB 4; Length 2148;

Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3657 ATGCAGTACTTAAGAAA 3674
|||||
Db 1700 ATGCAGTACTTAAGAAA 1683

RESULT 24

US-09-118-101A-3
; Sequence 3, Application US/08118101A
; Patent No. 5620892
; GENERAL INFORMATION:
; APPLICANT: Kuitz, Stephen E.
; APPLICANT: Knickerbocker, Aron M.
; APPLICANT: McCullough, John R.
; TITLE OF INVENTION: A STRAIN OF SACCCHAROMYCES CEREVISIAE
; TITLE OF INVENTION: EXPRESSING THE GENE ENCODING POTASSIUM TRANSPORTER MINK
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,101A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-5901
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2669
US-08-118-101A-3

Query Match 0.5%; Score 18; DB 1; Length 2669;

Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1376 TCTTTTGTGTGAGGAG 1393
|||||
Db 1147 TCTTTTGTGTGAGGAG 1164

RESULT 25

US-08-348-143-3
; Sequence 3, Application US/08348143
; Patent No. 5506205
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 5506205el Polypeptide of Protein p140 and DNAs
; TITLE OF INVENTION: encoding it
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,143
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
US-08-348-143-3

Query Match 0.5%; Score 18; DB 1; Length 4027;

Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTTGTGTTTCACT 91
|||||
Db 3523 ATTGGTTTGTGTTTCACT 3540

RESULT 26

US-08-348-143-4
; Sequence 4, Application US/08348143
; Patent No. 5506205
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 5506205el Polypeptide of Protein p140 and DNAs
; TITLE OF INVENTION: encoding it
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,143
; FILING DATE:
; CLASSIFICATION: 530

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103

INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 262..3243
; IDENTIFICATION METHOD: by similarity to some other pattern

US-08-348-143-4
Query Match 0.5%; Score 18; DB 1; Length 4027;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGCTTCAGT 91

Db 3523 ATTGGTTGCTTCAGT 3540

RESULT 27
US-08-571-785-3
; Sequence 3, Application US/08571785
; Patent No. 5804411
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 5804411el Polypeptide of Protein p140
; TITLE OF INVENTION: and DNAs encoding it
; NUMBER OF SEQUENCES: 16

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/571,785
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/348,143
; FILING DATE: 23-NOV-1994
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103

INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 262..3243
; IDENTIFICATION METHOD: by similarity to some other pattern

US-08-571-785-4
Query Match 0.5%; Score 18; DB 1; Length 4027;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGCTTCAGT 91

Db 3523 ATTGGTTGCTTCAGT 3540

RESULT 28
US-08-571-785-4
; Sequence 4, Application US/08571785
; Patent No. 5804411
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 5804411el Polypeptide of Protein p140
; TITLE OF INVENTION: and DNAs encoding it
; NUMBER OF SEQUENCES: 16

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/571,785

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

US-08-571-785-4

US-08-571-785-3

```
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/348,143
; FILING DATE: 23-NOV-1994
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 262..3243
; IDENTIFICATION METHOD: by similarity to some other pattern
US-08-571-785-4
```

Query Match 0.5%; Score 18; DB 1; Length 4027;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 74 ATTGGTGTGCTTCAGT 91
|||||
DB 3523 ATTGGTGTGCTTCAGT 3540
```

RESULT 29

```
US-09-192-435-3
; Sequence 3, Application US/09192435
; Patent No. 6303320
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 6303320el Polypeptide of Protein p140
; TITLE OF INVENTION: and DNAs encoding it
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,435
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/571,785
; FILING DATE: 13-DEC-1995
; APPLICATION NUMBER: 08/348,143
; FILING DATE: 23-NOV-1994
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
```

```
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
US-09-192-435-3
```

Query Match 0.5%; Score 18; DB 4; Length 4027;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 74 ATTGGTGTGCTTCAGT 91
|||||
DB 3523 ATTGGTGTGCTTCAGT 3540
```

RESULT 30

```
US-09-192-435-4
; Sequence 4, Application US/09192435
; Patent No. 6303320
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 6303320el Polypeptide of Protein p140
; TITLE OF INVENTION: and DNAs encoding it
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,435
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/571,785
; FILING DATE: 13-DEC-1995
; APPLICATION NUMBER: 08/348,143
; FILING DATE: 23-NOV-1994
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
```


TISSUE TYPE: skeletal muscle myoblast

CELL LINE: L6

FEATURE: CDS

LOCATION: 262..3243

IDENTIFICATION METHOD: by similarity to some other pattern

US-09-192-435-4

Query Match 0.5%; Score 18; DB 4; Length 4027;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGCTTCAGT 91

|||||

Db 3523 ATTGGTTGCTTCAGT 3540

RESULT 31

US-09-558-340-3

Sequence 3, Application US/09558340

Patent No. 6432913

GENERAL INFORMATION:

APPLICANT: TAJIMA, HISAO

APPLICANT: KITAGAWA, KOICHIRO

APPLICANT: OHNO, HIROYUKI

APPLICANT: UENO, TOSHIO

TITLE OF INVENTION: A No. 6432913el Polypeptide of Protein p140

TITLE OF INVENTION: and DNAs encoding it

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS

STREET: 2100 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20037-3202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/558,340

FILING DATE: 26-APR-2000

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/192,435

FILING DATE: 08-JAN-1998

APPLICATION NUMBER: 08/571,785

FILING DATE: 13-DEC-1995

APPLICATION NUMBER: 08/348,143

FILING DATE: 23-NOV-1994

APPLICATION NUMBER: JP 315806/1993

FILING DATE: 24-NOV-1993

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)293-7060

TELEFAX: (202)293-7860

TELEX: 6491103

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 4027 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORGANISM: rat

TISSUE TYPE: skeletal muscle myoblast

CELL LINE: L6

US-09-558-340-3

Query Match

Best Local Similarity 100.0%; Pred. No. 44;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGCTTCAGT 91

|||||

Db 3523 ATTGGTTGCTTCAGT 3540

RESULT 32

US-09-558-340-4

Sequence 4, Application US/09558340

Patent No. 6432913

GENERAL INFORMATION:

APPLICANT: TAJIMA, HISAO

APPLICANT: KITAGAWA, KOICHIRO

APPLICANT: OHNO, HIROYUKI

APPLICANT: UENO, TOSHIO

TITLE OF INVENTION: A No. 6432913el Polypeptide of Protein p140

TITLE OF INVENTION: and DNAs encoding it

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS

STREET: 2100 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20037-3202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/558,340

FILING DATE: 26-APR-2000

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/192,435

FILING DATE: 08-JAN-1998

APPLICATION NUMBER: 08/571,785

FILING DATE: 13-DEC-1995

APPLICATION NUMBER: 08/348,143

FILING DATE: 23-NOV-1994

APPLICATION NUMBER: JP 315806/1993

FILING DATE: 24-NOV-1993

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)293-7060

TELEFAX: (202)293-7860

TELEX: 6491103

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 4027 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORGANISM: rat

TISSUE TYPE: skeletal muscle myoblast

CELL LINE: L6

FEATURE:

NAME/KEY: CDS

LOCATION: 262..3243

IDENTIFICATION METHOD: by similarity to some other pattern

US-09-558-340-4

Query Match

Best Local Similarity 100.0%; Pred. No. 44;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGCTTCAGT 91

|||||

Db 3523 ATTGGTTGCTTCAGT 3540

RESULT 33

US-08-915-868-1
; Sequence 1, Application US/08915868
; Patent No. 5314233
; GENERAL INFORMATION:
; APPLICANT: Mundy, Gregory R.
; TITLE OF INVENTION: SCREENING ASSAY FOR THE IDENTIFICATION
; TITLE OF INVENTION: OF AGENTS WHICH INHIBIT CANCER
; TITLE OF INVENTION: METASTASIS TO BONE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,868
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025,215
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: OSTs:002P21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4348 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-915-868-1

Query Match 0.5%; Score 18; DB 2; Length 4348;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 576 AGAAATTTTTCCTTT 593
|||||
Db 4180 AGAAATTTTTCCTTT 4197

RESULT 34

US-09-315-793-11
; Sequence 11, Application US/09315793
; Patent No. 6221597
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-048
; CURRENT APPLICATION NUMBER: US/09/315,793
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 11
; LENGTH: 5097
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae

US-09-315-793-11

Query Match 0.5%; Score 18; DB 4; Length 5097;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 873 AGGAAAGTTTTTCCACT 890
|||||
Db 2560 AGGAAAGTTTTTCCACT 2577
RESULT 35
US-09-058-389A-5/c
; Sequence 5, Application US/09058389A
; Patent No. 6130065
; GENERAL INFORMATION:
; APPLICANT: Belt, Judith A.
; APPLICANT: Crawford, Charles R.
; TITLE OF INVENTION: A NITROBENZYL MERCAPTOPURINE RIBOSIDE
; TITLE OF INVENTION: (NBMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,389A
; FILING DATE: April 9, 1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-013N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6354 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
US-09-058-389A-5

Query Match 0.5%; Score 18; DB 3; Length 6354;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2068 GGAAGCTGTAGATCAGCT 2085
|||||
Db 6320 GGAAGCTGTAGATCAGCT 6303

RESULT 36

US-09-611-781-5/c
; Sequence 5, Application US/09611781
; Patent No. 6423829
; GENERAL INFORMATION:
; APPLICANT: Belt, Judith A.

```
; APPLICANT: Crawford, Charles R.
; APPLICANT: Patel, Divyen
; TITLE OF INVENTION: A NITROBENZYL MERCAPTOPURINE RIBOSIDE
; TITLE OF INVENTION: (NMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/611,781
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/058,389
; FILING DATE: April 9, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-013N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6354 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; US-09-611-781-5

Query Match 0.5%; Score 18; DB 4; Length 6354;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2068 GGAAGCTGTAGATCAGCT 2085
Db 6320 GGAAGCTGTAGATCAGCT 6303
|||||

RESULT 37
US-09-227-219-130/C
; Sequence 130, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisa
; APPLICANT: Katsuki, Pamela J.
; APPLICANT: Furuta, Hiroto
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; TITLE OF INVENTION: AND HNF-4ALPHA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
```

```
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,219
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,679
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,056
; FILING DATE: 02-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,719
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10014 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-927-219-130

Query Match 0.5%; Score 18; DB 4; Length 10014;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1384 TGATGAGGAGTGGGAGAA 1401
Db 4194 TGATGAGGAGTGGGAGAA 4177
|||||

RESULT 38
US-08-991-840A-1
; Sequence 1, Application US/08991840A
; Patent No. 6261570
; GENERAL INFORMATION:
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Bruce Crise
; APPLICANT: Mark Steve Oberste
; APPLICANT: Shannon Schmura
; TITLE OF INVENTION: Live Attenuated Virus Vaccines for Eastern Equine Encephali
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Charles H. Harris
; STREET: USA MRMC - MRMC-JA
; CITY: FORT DETRICK, FREDERICK
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21702-5012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,840A
; FILING DATE: December 16, 1997
; CLASSIFICATION: 424
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: Provisional Application 60/047162,
; FILING DATE: May 20, 1997
; APPLICATION NUMBER: Provisional Application 60/053,652
; FILING DATE: July 24, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles H. Harris
; REGISTRATION NUMBER: 34,616
; REFERENCE/DOCKET NUMBER: 003/058/SAP RIID 96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 619-2065
; TELEFAX: (301) 619-5034
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11492 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; FEATURE: OTHER INFORMATION: N at all occurrences is = unknown.
; FEATURE: OTHER INFORMATION: K at all
; FEATURE: occurrences is = G or T
US-08-991-840A-1

Query Match      0.5%; Score 18; DB 4; Length 11492;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 GTGCTGACAAATCATCA 380
|||||
Db 2636 GTGCTGACAAATCATCA 2653

RESULT 39
US-09-819-993-3/c
; Sequence 3, Application us/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; MOLECULE TYPE: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(28001)
; OTHER INFORMATION: n = A,T,C or G
US-09-819-993-3

Query Match      0.5%; Score 18; DB 4; Length 28001;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3409 TTTCATTAGCTTCCCTG 3426
|||||
Db 19180 TTTCATTAGCTTCCCTG 19163

RESULT 40
US-09-215-694-19/c
; Sequence 19, Application US/09215694B
; Patent No. 6391583
; GENERAL INFORMATION:
; APPLICANT: Wisconsin Alumni Research Foundation
; APPLICANT: Hutchinson, Charles R.
; APPLICANT: Kennedy, Jonathan n.m.i.
```

```
; APPLICANT: Park, Cheonseok n.m.i
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
; FILE REFERENCE: 960296.95718
; CURRENT APPLICATION NUMBER: US/09/215,694B
; CURRENT FILING DATE: 1999-12-18
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 19
; LENGTH: 31328
; TYPE: DNA
; ORGANISM: Aspergillus terreus
US-09-215-694-19

Query Match      0.5%; Score 18; DB 4; Length 31328;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2048 GTTCTGACAGAGCAAA 2065
|||||
Db 27833 GTTCTGACAGAGCAAA 27816

RESULT 41
US-09-453-702B-116/c
; Sequence 116, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 116:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45175
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 116:
US-09-453-702B-116

Query Match      0.5%; Score 18; DB 4; Length 45175;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 2270 AGAATATTATTCACAAAT 2287
|||||
Db 56 AGAATATTATTCACAAAT 39

RESULT 42

US-09-188-930-206
; Sequence 206, Application US/09:88930A
; Patent No. 6150502
; GENERAL INFORMATION:

; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCES: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09

; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 206
; LENGTH: 480

; TYPE: DNA

; ORGANISM: Mouse

US-09-188-930-206

Query Match 0.5%; Score 17; DB 3; Length 480;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 TGCTTCAGTTTCAGTCA 99
|||||
Db 262 TGCTTCAGTTTCAGTCA 278

RESULT 43

US-08-460-739-2
; Sequence 2, Application US/08460739
; Patent No. 5698416
; GENERAL INFORMATION:

; APPLICANT: Wolf, Marcia K.
; APPLICANT: Cassels, Frederick J.
; APPLICANT: Bell, Brian A.

; TITLE OF INVENTION: Improved Methods for Production of
Antigens Under Control of Tempature-Regulated Promotors

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Glenna Hendricks
; STREET: 9669 A Main Street
; CITY: Fairfax

; STATE: VA

; COUNTRY: US

; ZIP: 22031

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,739

; FILING DATE: 20-APR-1995

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Hendricks, Glenna M.

; REGISTRATION NUMBER: 32,535

; REFERENCE/DOCKET NUMBER: wolf2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 425-4250

; TELEFAX: (703) 425-2767

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: DNA for encoding CFAL protein
US-08-460-739-2

Query Match 0.5%; Score 17; DB 1; Length 512;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3212 GCATCAAAACTTTTGA 3228
|||||
Db 174 GCATCAAAACTTTTGA 190

RESULT 44

US-09-134-001C-1886/c
; Sequence 1886, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 1886

; LENGTH: 777

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1886

Query Match 0.5%; Score 17; DB 4; Length 777;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 924 CTGGTTTCACCTTTTC 940
|||||
Db 439 CTGGTTTCACCTTTTC 423

RESULT 45

US-08-961-527-264/c
; Sequence 264, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 264:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 795 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-264

Query Match 0.5%; Score 17; DB 4; Length 795;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2865 ACTTCCAAAGACTGGT 2881
|||||
DB 655 ACTTCCAAAGACTGGT 639

RESULT 46

; Sequence 10, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:

; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alex
; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; FILE REFERENCE: L0461/7051
; CURRENT APPLICATION NUMBER: US/09/166,350A
; EARLIER FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: US 09/166,350
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-166-350-10

Query Match 0.5%; Score 17; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 CTGTTATAGCTCTTTA 758
|||||
DB 407 CTGTTATAGCTCTTTA 391

RESULT 47

; Sequence 36, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew

; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 896
; TYPE: DNA
; ORGANISM: mouse
US-09-188-930-36

Query Match 0.5%; Score 17; DB 3; Length 896;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 TGCTTCAGTTTCAGTCA 99
|||||
DB 194 TGCTTCAGTTTCAGTCA 210

RESULT 48

; Sequence 7, Application US/08671525B
; Patent No. 5703220
; GENERAL INFORMATION:
; APPLICANT: Yamada, Tadataka
; APPLICANT: Gantz, Ira
; TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: US
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/671,525B
; FILING DATE: June 27, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-000853DVB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810)641-1600
; TELEFAX: (810)641-0270
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..996
US-08-671-525B-7

Query Match 0.5%; Score 17; DB 1; Length 996;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 TTTCATTCAGTGGAC 2866
|||||
Db 422 TTTCATTCAGTGGAC 438

RESULT 49
US-08-672-109B-7
; Sequence 7, Application US/08672109B
; Patent No. 5710265
; GENERAL INFORMATION:
; APPLICANT: Yamada, Tadataka
; APPLICANT: Gantz, Ira
; TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: US
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672,109B
; FILING DATE: June 27, 1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-000853DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810)641-1600
; TELEFAX: (810)641-0270
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..996
US-08-672-109B-7

Query Match 0.5%; Score 17; DB 1; Length 996;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 TTTCATTCAGTGGAC 2866
|||||
Db 422 TTTCATTCAGTGGAC 438

RESULT 50
US-08-842-045-7
; Sequence 7, Application US/08842045
; Patent No. 5817787
; GENERAL INFORMATION:
; APPLICANT: Yamada, Tadataka
; APPLICANT: Gantz, Ira
; TITLE OF INVENTION: Genes Encoding Melanocortin Receptors

; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: US
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/842,045
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-000853DVE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810)641-1600
; TELEFAX: (810)641-0270
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..996
US-08-842-045-7

Query Match 0.5%; Score 17; DB 1; Length 996;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 TTTCATTCAGTGGAC 2866
|||||
Db 422 TTTCATTCAGTGGAC 438

Search completed: November 6, 2002, 03:36:45
Job time : 687 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 23:31:02 ; Search time 158 Seconds
(without alignments)
7867.093 Million cell updates/sec

Title: US-09-782-874-1
Perfect score: 3731
Sequence: 1 GAATAATCTTTACTTACTT.....AGTTTCATCTTCTCTCTAAA 3731

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 310279 seqs, 166577418 residues

Word size : 0

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3731	100.0	3731	10	US-09-782-874-1
2	25	0.7	25	10	US-09-782-874-4
3	25	0.7	25	10	US-09-782-874-6
4	24	0.6	25	10	US-09-782-874-5
5	20	0.5	398	10	US-09-998-598-194
6	20	0.5	407	10	US-09-962-436-82
7	20	0.5	409	10	US-09-998-598-772
8	20	0.5	413	10	US-09-998-598-2240
9	20	0.5	1750	10	US-09-840-787-54
10	19	0.5	34	10	US-09-782-874-7
11	19	0.5	267	10	US-09-563-817-946
12	19	0.5	461	10	US-09-864-761-11608
13	19	0.5	4656	10	US-09-764-877-2690
14	18	0.5	35	10	US-09-782-874-9
15	18	0.5	256	10	US-09-923-876-1838
16	18	0.5	358	10	US-09-960-352-1347
17	18	0.5	385	10	US-09-867-701-4817
18	18	0.5	474	10	US-09-864-761-11283
19	18	0.5	500	10	US-09-783-590-8123

c 20	18	0.5	1016	10	US-09-925-239-176	Sequence 176, App
c 21	18	0.5	1467	10	US-09-815-242-7347	Sequence 7347, App
c 22	18	0.5	1497	10	US-09-887-576-629	Sequence 629, App
c 23	18	0.5	1944	10	US-09-887-576-619	Sequence 619, App
c 24	18	0.5	2148	10	US-09-841-132-328	Sequence 328, App
c 25	18	0.5	3048	12	US-10-007-693-80	Sequence 80, Appl
c 26	18	0.5	3760	10	US-09-822-849A-162	Sequence 162, App
c 27	18	0.5	3781	12	US-10-052-586-453	Sequence 453, App
c 28	18	0.5	4348	10	US-09-879-445-1	Sequence 1, Appl
c 29	18	0.5	32192	10	US-09-764-877-3657	Sequence 3657, App
c 30	18	0.5	49744	10	US-09-927-091-4	Sequence 4, Appl
c 31	18	0.5	170834	10	US-09-835-232-7	Sequence 7, Appl
c 32	17	0.5	150	10	US-09-923-876-3754	Sequence 3754, App
c 33	17	0.5	213	10	US-09-878-574-9084	Sequence 9084, App
c 34	17	0.5	258	10	US-09-878-574-7593	Sequence 7593, App
c 35	17	0.5	269	10	US-09-969-373-41	Sequence 41, Appl
c 36	17	0.5	274	10	US-09-294-093B-1096	Sequence 1096, App
c 37	17	0.5	308	10	US-09-864-761-31520	Sequence 31520, App
c 38	17	0.5	329	10	US-09-880-107-706	Sequence 706, App
c 39	17	0.5	353	10	US-09-917-800A-623	Sequence 623, App
c 40	17	0.5	368	10	US-09-864-761-13956	Sequence 13956, App
c 41	17	0.5	387	10	US-09-974-300-7289	Sequence 7289, App
c 42	17	0.5	391	10	US-09-864-761-30520	Sequence 30520, App
c 43	17	0.5	391	10	US-09-563-817-878	Sequence 878, App
c 44	17	0.5	406	10	US-09-983-965-401	Sequence 401, App
c 45	17	0.5	411	10	US-09-867-550-1095	Sequence 1095, App
c 46	17	0.5	419	10	US-09-960-352-10898	Sequence 10898, App
c 47	17	0.5	421	10	US-09-880-107-2705	Sequence 2705, App
c 48	17	0.5	430	10	US-09-864-761-14986	Sequence 14986, App
c 49	17	0.5	436	10	US-09-864-761-2826	Sequence 2826, App
c 50	17	0.5	436	10	US-09-983-965-2934	Sequence 2934, App

ALIGNMENTS

RESULT 1
US-09-782-874-1
; Sequence 1, Application US/09782874
; Patent No. US20010023067A1

GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDRP)

NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/782.874
FILING DATE: 08-Feb-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/811,583

FILING DATE: 05-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Haley, James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: MPG-1

TELECOMMUNICATION INFORMATION:


```

, TELEPHONE: 212-596-9000
,
, TELEFAX: 212-536-9090
,
, INFORMATION FOR SEQ ID NO: 1:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 3731 base pairs
,
, TYPE: nucleic acid
,
, STRANDEDNESS: single
,
, TOPOLOGY: linear
,
, MOLECULE TYPE: cDNA
,
, ORIGINAL SOURCE:
,
, ORGANISM: Tomato
,
, FEATURE:
,
, NAME/KEY: CDS
,
, LOCATION: 194..3535
,
, SEQUENCE DESCRIPTION: SEQ ID
US-09-782-874-1

```

	Query Match	100.0%; Score 3731; DB 10; Length 3731;	
	Best Local Similarity	100.0%; Pred. No. 0;	
	Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	GAAATATTCTTTACTACTTTCACAGGGATGACTCATCACTCCCTCCAAGTCCTTGCTGT	60
DB	1		
QY	61	GTGTGATAATAAATTTGGTGTCTCTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT	120
DB	61		
QY	121	TTTGCATAACTTCAGGGGGTYATTCAGTGGTGTAGCATTTGAAGTGAAGTCACTGCACTT	180
DB	121	TTTGCATAACTTCAGGGGGTATTCAGTGGTGTAGCATTTGAAGTGAAGTCACTGCACTT	180
QY	181	GGAAATTTGGCTACATGTGGAAAGACAATTCAGTTTTCGGATTCCTTATCTCTCTCTGC	240
DB	181	GGAAATTTGGCTACATGTGGAAAGACAATTCAGTTTTCGGATTCCTTATCTCTCTCTGC	240
QY	241	GGAGTGGTTAAGTCATTCTTTAGAGAATAATACAGGATATGGAACGTGTATGTCATTTGA	300
DB	241	GGAGTGGTTAAGTCATTCTTTAGAGAATAATACAGGATATGGAACGTGTATGTCATTTGA	300
QY	301	GGTTAAACAGTCCCAAGCAGGATCTAGAGCATTTGCCAAGTTCAATTTGCCGACAACAT	360
DB	301	GGTTAAACAGTCCCAAGCAGGATCTAGAGCATTTGCCAAGTTCAATTTGCCGACAACAT	360
QY	361	AAGTGTGTGACAAAATCATCACTTTGGCTAATAACAGGCTGTATTTGGTCTCTCTTATTT	420
DB	361	AAGTGTGTGACAAAATCATCACTTTGGCTAATAACAGGCTGTATTTGGTCTCTCTTATTT	420
QY	421	GAAGGCTTGGGAAATGAAAACGTGATNTGCCAAGTCCGGGCATATGTGGATCAGATGA	480
DB	421	GAAGGCTTGGGAAATGAAAACGTGATNTGCCAAGTCCGGGCATATGTGGATCAGATGA	480
QY	481	TGGCATTAACCTTTGAATTTCCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG	540
DB	481	TGGCATTAACCTTTGAATTTCCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG	540
QY	541	TACAGAAGTTTCAATTCAAATTTGGCATTTGGATGGAAGAAATTTTTTTCTCTTTATCTAG	600
DB	541	TACAGAAGTTTCAATTCAAATTTGGCATTTGGATGGAAGAAATTTTTTTCTCTTTATCTAG	600
QY	601	TGCTTCAGTGTACTATAAATTCAGCTTTCATATGAAAAATATATGCGAGGTTGTCTCCA	660
DB	601	TGCTTCAGTGTACTATAAATTCAGCTTTCATATGAAAAATATATGCGAGGTTGTCTCCA	660
QY	661	TGCTCCATATGCTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTGCCTCCGAT	720
DB	661	TGCTCCATATGCTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTGCCTCCGAT	720
QY	721	CTATAAGAGACTTGAAGAACCTCCTGTTATAGTCTCTTTAAGGAAACTCCTGATGATCATGT	780
DB	721	CTATAAGAGACTTGAAGAACCTCCTGTTATAGTCTCTTTAAGGAAACTCCTGATGATCATGT	780

Qy	781	GGTGAGGACAACAGATTTCCCTCCAFCTTGTGGATAGGGCTATCTCTTACGTTATGTATTGCA	840
Db	781	GGTGAGGACAACAGATTTCCCTCCAFCTTGTGGATAGGGCTATCTCTTACGTTATGTATTGCA	840
Qy	841	GTTCCGTFAGGGTGTTCGCTCTCCAATTTCCAGGAAGAAGTTTTTTCCTACTATGCAGAAAGC	900
Db	841	GTTCCGTFAGGGTGTTCGCTCTCCAATTTCCAGGAAGAAGTTTTTTCCTACTATGCAGAAAGC	900
Qy	901	TGAAAACAATATTACTTTTACAGACTGGTTTTCACCTTTTTCGFTCTCTCAAAAATCGGCTCT	960
Db	901	TGAAAACAATATTACTTTTACAGACTGGTTTTCACCTTTTTCGFTCTCTCAAAAATCGGCTCT	960
Qy	961	GGTTCOCAATGTCACGCCCTCGGGAAGGAATTTCAATTCCTACAGATTTTGTGTCAAAT	1020
Db	961	GGTTCOCAATGTCACGCCCTCGGGAAGGAATTTCAATTCCTACAGATTTTGTGTCAAAT	1020
Qy	1021	TAGTCTTTGGTACAGCATGGATGTCATACTCGGCCAGCATTAAATGTCTACTTTTTTCCG	1080
Db	1021	TAGTCTTTGGTACAGCATGGATGTCATACTCGGCCAGCATTAAATGTCTACTTTTTTCCG	1080
Qy	1081	ATTAGTTGATCCFCGAGGAGAAATGTGCATCGCATGGCATGGCATGGCATGGCATGGCATGGTA	1140
Db	1081	ATTAGTTGATCCFCGAGGAGAAATGTGCATCGCATGGCATGGCATGGCATGGCATGGCATGGTA	1140
Qy	1141	CTATATAAAGAGTGTGTATTGATCCCGTAGGTGGCTCACTGAGCAGTAGTATGGSTA	1200
Db	1141	CTATATAAAGAGTGTGTATTGATCCCGTAGGTGGCTCACTGAGCAGTAGTATGGSTA	1200
Qy	1201	TCTCAAGGTFAGACAACCTTCCAAAATCTCCGTCCTCACTTTAGATNATGGTGTGGTGA	1260
Db	1201	TCTCAAGGTFAGACAACCTTCCAAAATCTCCGTCCTCACTTTAGATNATGGTGTGGTGA	1260
Qy	1261	TGTAAGAAGGTOCTTAGTAACACCATGCAAAAGTTATTTTGTGTCTCCAGAGGTTAATGT	1320
Db	1261	TGTAAGAAGGTOCTTAGTAACACCATGCAAAAGTTATTTTGTGTCTCCAGAGGTTAATGT	1320
Qy	1321	TTCCAAATCGGGTCTCCGCAATATTCTGAAGACATAGATAAATCTCTCTCGTGTCTCTTT	1380
Db	1321	TTCCAAATCGGGTCTCCGCAATATTCTGAAGACATAGATAAATCTCTCTCGTGTCTCTTT	1380
Qy	1381	TGTTGATGAGAGTGGGAGAACTGTATTCTACAGACTTATACCAAAGCAAGTACTGG	1440
Db	1381	TGTTGATGAGAGTGGGAGAACTGTATTCTACAGACTTATACCAAAGCAAGTACTGG	1440
Qy	1441	AAGTGTGTCCAGAACAACTATGAGAGAGATCTTATCAACTCTGCGGAAAGCCTTTGT	1500
Db	1441	AAGTGTGTCCAGAACAACTATGAGAGAGATCTTATCAACTCTGCGGAAAGCCTTTGT	1500
Qy	1501	AATTGTGTATAAAAATTTGAATTTCTTGCAATTTTCATCGAGCCAGTTGCGSGGATAATC	1560
Db	1501	AATTGTGTATAAAAATTTGAATTTCTTGCAATTTTCATCGAGCCAGTTGCGSGGATAATC	1560
Qy	1561	AGTGTGGATGTTTGATCAAGACCTGGCCTTACTGCCAAATGATATAAGAGCTTGGATGGG	1620
Db	1561	AGTGTGGATGTTTGATCAAGACCTGGCCTTACTGCCAAATGATATAAGAGCTTGGATGGG	1620
Qy	1621	TGATTTTTCGCAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTTGG	1680
Db	1621	TGATTTTTCGCAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTTGG	1680
Qy	1681	TTCTCCAGAGAGACTTTCAGTGTCTTAGGCATGAGATTTGAAGTTATTCCCGATGTAAA	1740
Db	1681	TTCTCCAGAGAGACTTTCAGTGTCTTAGGCATGAGATTTGAAGTTATTCCCGATGTAAA	1740
Qy	1741	GGTTCATGGAACCGATGTCTTTTCTGTATGGAAATCGTAAANATATCTGGTGACTTTGC	1800
Db	1741	GGTTCATGGAACCGATGTCTTTTCTGTATGGAAATCGTAAANATATCTGGTGACTTTGC	1800
Qy	1801	TCATAGAGTTGCCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGAGTTCGTTA	1860
Db	1801	TCATAGAGTTGCCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGAGTTCGTTA	1860
Qy	1861	TGSGTGATATAAGGTTGTGGGTGTGATCCGGATTCATCAATAGAGTTGTCTTTGAG	1920

1861 TGGTGATATAAGGTGTGTGGTGTGATCCGGATTCAATGAATGCTTTGAG 1920
1921 AAAGAGCATGTCGAAATGATGATCAGACACATCAATGAATGATGCTGATGAGCAA 1980
1921 AAAGAGCATGTCGAAATGATGATCAGACACATCAATGAATGATGCTGATGAGCAA 1980
1981 ATATCAGCCTTGTTATCTTAATCTCAACTGATTACCGCTCTGCTACACTGGAGTGA 2040
1981 ATATCAGCCTTGTTATCTTAATCTCAACTGATTACCGCTCTGCTACACTGGAGTGA 2040
2041 AGATGAAGTCTCCAGACAGACAGCAAGAGCTGTAGATCAGCTTATGCTATCTTGA 2100
2041 AGATGAAGTCTCCAGACAGACAGCAAGAGCTGTAGATCAGCTTATGCTATCTTGA 2100
2101 TGATTCCTTTGAAGCAGCAGGAGGCTTTGGAATGATGCTCTCCAGAGACACATATAT 2160
2101 TGATTCCTTTGAAGCAGCAGGAGGCTTTGGAATGATGCTCTCCAGAGACACATATAT 2160
2161 TCTCAAGGCAATGCTAACTGTGTTATAGCCTGATGCTGAGCCCTTCTTTCAATGAT 2220
2161 TCTCAAGGCAATGCTAACTGTGTTATAGCCTGATGCTGAGCCCTTCTTTCAATGAT 2220
2221 GTTCCAAACCTTCGCGCATCCAAAGTTCGCTGATTTGCGGACTAGATCAAGAATATTAT 2280
2221 GTTCCAAACCTTCGCGCATCCAAAGTTCGCTGATTTGCGGACTAGATCAAGAATATTAT 2280
2281 TCCAAATGGAAGACAAATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATGCTCA 2340
2281 TCCAAATGGAAGACAAATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATGCTCA 2340
2341 GGTGTTTGTTCAGTTACTGTGCTGACATGGAGAGCTTTCTGACGATTTACATCAATT 2400
2341 GGTGTTTGTTCAGTTACTGTGCTGACATGGAGAGCTTTCTGACGATTTACATCAATT 2400
2401 TAATAACAGCAGATCCCAACAGTAATTTCACTCTCAAGGAAATGCTGTTGTTGCAA 2460
2401 TAATAACAGCAGATCCCAACAGTAATTTCACTCTCAAGGAAATGCTGTTGTTGCAA 2460
2461 AAATCCCATGCTGATCCTGCTGATATCTGTTTAAAGCGTGTAATGTTGAGCGCT 2520
2461 AAATCCCATGCTGATCCTGCTGATATCTGTTTAAAGCGTGTAATGTTGAGCGCT 2520
2521 GCACCATAGTATGATGTTGTTATCCCTCAGAAAGAAAGAGCTCATCCGAATGA 2580
2521 GCACCATAGTATGATGTTGTTATCCCTCAGAAAGAAAGAGCTCATCCGAATGA 2580
2581 ATGTTCTGGAGTATGTTGATGGGATATCTACTTTGTTGCTGGATCAAGACATGAT 2640
2581 ATGTTCTGGAGTATGTTGATGGGATATCTACTTTGTTGCTGGATCAAGACATGAT 2640
2641 CCCGCCAAGGCAAGTCCAGCGGATGGAATATCCCTCAGCAGCAGCAGATAGTTGGACCA 2700
2641 CCCGCCAAGGCAAGTCCAGCGGATGGAATATCCCTCAGCAGCAGCAGATAGTTGGACCA 2700
2701 TGATGTCACAAATTGAGGAATGGAAGTACTTCAACAACTATATGTTGAATGACAGTTT 2760
2701 TGATGTCACAAATTGAGGAATGGAAGTACTTCAACAACTATATGTTGAATGACAGTTT 2760
2761 GGGATCATAGCAATCCCATGCTGCTGATTTGAGACAGAGACCTGATATGCCCATGAG 2820
2761 GGGATCATAGCAATCCCATGCTGCTGATTTGAGACAGAGACCTGATATGCCCATGAG 2820
2821 TGATCCATGCAAAAACCTTGTGAGCTCTTTTCAATTCAGTGGAGCTTTCCAAAGACTGG 2880
2821 TGATCCATGCAAAAACCTTGTGAGCTCTTTTCAATTCAGTGGAGCTTTCCAAAGACTGG 2880
2881 TGTTCCTGCTGAATACATCTAGTTGCGCCCTTAAGAAATACCCAGACTTCATGATATA 2940
2881 TGTTCCTGCTGAATACATCTAGTTGCGCCCTTAAGAAATACCCAGACTTCATGATATA 2940
2941 GCCGGACAAGACCAAGCTATATCTCAGAAAGAGTTATGGAAGCTTTTCAGGAAAGTGA 3000

2941 GCCGGACAAGACCAAGCTATATCTCAGAAAGAGTTATTGGAAGAGCTTTTCAGGAAAGTGA 3000
3001 GGACAAAGCACCTCAGCTAGCTCTATCCGACCTTCAACAAGAGATGTTGCAAGGAGATC 3060
3001 GGACAAAGCACCTCAGCTAGCTCTATCCGACCTTCAACAAGAGATGTTGCAAGGAGATC 3060
3061 ATATGATGCTGATATGGAAGTGTGATGGAATTTAAGATTTACATTGACGAAAGCTTTTACATA 3120
3061 ATATGATGCTGATATGGAAGTGTGATGGAATTTAAGATTTACATTGACGAAAGCTTTTACATA 3120
3121 CAAACTGATATGACAAAGCTGGTAAATTTAATGGACTACTATGCGATATAAAACAGA 3180
3121 CAAACTGATATGACAAAGCTGGTAAATTTAATGGACTACTATGCGATATAAAACAGA 3180
3181 GGTGAAATACTTGTGTTGGCATTTAGAGGCATCAAAAACCTTTGACCGCAAGAAAGA 3240
3181 GGTGAAATACTTGTGTTGGCATTTAGAGGCATCAAAAACCTTTGACCGCAAGAAAGA 3240
3241 TGTGAGGCGCATTTAGTGTGCTGTGAGGCGCTTTGAGAGGAGGCAAGAGCCTGGTTCAA 3300
3241 TGTGAGGCGCATTTAGTGTGCTGTGAGGCGCTTTGAGAGGAGGCAAGAGCCTGGTTCAA 3300
3301 GAGGCGCTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACCATACATA 3360
3301 GAGGCGCTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACCATACATA 3360
3361 TCATCTACATATTTGGGTTGCTTACAAATCAGGGGTTGAAAAGAGCTCATTTCAATAGCTT 3420
3361 TCATCTACATATTTGGGTTGCTTACAAATCAGGGGTTGAAAAGAGCTCATTTCAATAGCTT 3420
3421 TCCTCGTGTGTTTATGACCACTAATCCAGATTAAGAGGCAAGAGCAGTAACAGGCC 3480
3421 TCCTCGTGTGTTTATGACCACTAATCCAGATTAAGAGGCAAGAGCAGTAACAGGCC 3480
3481 AGTCTCAACTTGTCTATCTCTCAGGCGCTCACTAGTACAGATTAAGTGTGAAATGAGA 3540
3481 AGTCTCAACTTGTCTATCTCTCAGGCGCTCACTAGTACAGATTAAGTGTGAAATGAGA 3540
3541 TTCCAGTCGAGCGTTAAGCTGATATATATATATGTAATAGGCTGTGATCAAGAAAC 3600
3541 TTCCAGTCGAGCGTTAAGCTGATATATATATATGTAATAGGCTGTGATCAAGAAAC 3600
3601 TGTATGCAATTTGACTACCTTTTCTCTTTAAAAGTGCATGAAGCTGCAACATATATGC 3660
3601 TGTATGCAATTTGACTACCTTTTCTCTTTAAAAGTGCATGAAGCTGCAACATATATGC 3660
3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTAATATGATGATTTGATTTGATTCATC 3720
3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTAATATGATGATTTGATTTGATTCATC 3720
3721 TTTCTTCTAAA 3731
3721 TTTCTTCTAAA 3731

RESULT 2

US-09-782-874-4/c
; Sequence 4, Application US/09782874
; Patent No. US20010023067A1

GENERAL INFORMATION:

APPLICANT: Wassenecker, Michael
Riedel, Leonhard
Schiebel, Winfried
Sanger, Heinz

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
RNA-DIRECTED RNA POLYMERASE (RdRp)

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York

```
;
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-782-874-4
Query Match 0.7%; Score 25; DB 10; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1838 CCATCTGCTTCCAGATTCGTTATG 1862
; DB 25 CCATCTGCTTCCAGATTCGTTATG 1
;
; RESULT 3
; US-09-782-874-6/c
; Sequence 6, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wasseneegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
;
; US-09-782-874-5/c
; Sequence 5, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wasseneegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-782-874-6
Query Match 0.7%; Score 25; DB 10; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2659 GCCGATGGAATATCTCCAGCACCC 2683
; DB 25 GCCGATGGAATATCTCCAGCACCC 1
;
; RESULT 4
; US-09-782-874-5/c
; Sequence 5, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wasseneegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
```

Matches	20;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935						
Db	384	TTGAGAAAGAGCATGTCGAA	365						
<p>RESULT 7</p> <p>US-09-998-598-772/c</p> <p>; Sequence 772, Application US/09998598</p> <p>; Patent No. US20020150922A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Stolk, John A.</p> <p>; APPLICANT: Xu, Jiangchun</p> <p>; APPLICANT: Chenault, Ruth A.</p> <p>; APPLICANT: Meagher, Madelein Joy</p> <p>; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND</p> <p>; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER</p> <p>; FILE REFERENCE: 210121.561</p> <p>; CURRENT APPLICATION NUMBER: US/09/998,598</p> <p>; CURRENT FILING DATE: 2001-11-16</p> <p>; NUMBER OF SEQ ID NOS: 2606</p> <p>; SOFTWARE: Corixa Invention Disclosure Database</p> <p>; SEQ ID NO: 772</p> <p>; LENGTH: 409</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Homo sapiens</p> <p>US-09-998-598-772</p> <p>Query Match 0.5%; Score 20; DB 10; Length 409;</p> <p>Best Local Similarity 100.0%; Pred. No. 2.7;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>									
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935						
Db	360	TTGAGAAAGAGCATGTCGAA	341						
<p>RESULT 8</p> <p>US-09-998-598-2240/c</p> <p>; Sequence 2240, Application US/09998598</p> <p>; Patent No. US20020150922A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Stolk, John A.</p> <p>; APPLICANT: Xu, Jiangchun</p> <p>; APPLICANT: Chenault, Ruth A.</p> <p>; APPLICANT: Meagher, Madelein Joy</p> <p>; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND</p> <p>; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER</p> <p>; FILE REFERENCE: 210121.561</p> <p>; CURRENT APPLICATION NUMBER: US/09/998,598</p> <p>; CURRENT FILING DATE: 2001-11-16</p> <p>; NUMBER OF SEQ ID NOS: 2606</p> <p>; SOFTWARE: Corixa Invention Disclosure Database</p> <p>; SEQ ID NO: 2240</p> <p>; LENGTH: 413</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Homo sapiens</p> <p>US-09-998-598-2240</p> <p>Query Match 0.5%; Score 20; DB 10; Length 413;</p> <p>Best Local Similarity 100.0%; Pred. No. 2.8;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>									
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935						
Db	364	TTGAGAAAGAGCATGTCGAA	345						
<p>RESULT 9</p> <p>US-09-840-787-54</p> <p>; Sequence 54, Application US/09840787</p> <p>; Patent No. US20020058264A1</p>									

Matches	20;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935						
Db	384	TTGAGAAAGAGCATGTCGAA	365						
<p>RESULT 7</p> <p>US-09-998-598-772/c</p> <p>; Sequence 772, Application US/09998598</p> <p>; Patent No. US20020150922A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Stolk, John A.</p> <p>; APPLICANT: Xu, Jiangchun</p> <p>; APPLICANT: Chenault, Ruth A.</p> <p>; APPLICANT: Meagher, Madelein Joy</p> <p>; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND</p> <p>; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER</p> <p>; FILE REFERENCE: 210121.561</p> <p>; CURRENT APPLICATION NUMBER: US/09/998,598</p> <p>; CURRENT FILING DATE: 2001-11-16</p> <p>; NUMBER OF SEQ ID NOS: 2606</p> <p>; SOFTWARE: Corixa Invention Disclosure Database</p> <p>; SEQ ID NO: 772</p> <p>; LENGTH: 409</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Homo sapiens</p> <p>US-09-998-598-772</p> <p>Query Match 0.5%; Score 20; DB 10; Length 409;</p> <p>Best Local Similarity 100.0%; Pred. No. 2.7;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>									
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935						
Db	360	TTGAGAAAGAGCATGTCGAA	341						
<p>RESULT 8</p> <p>US-09-998-598-2240/c</p> <p>; Sequence 2240, Application US/09998598</p> <p>; Patent No. US20020150922A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Stolk, John A.</p> <p>; APPLICANT: Xu, Jiangchun</p> <p>; APPLICANT: Chenault, Ruth A.</p> <p>; APPLICANT: Meagher, Madelein Joy</p> <p>; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND</p> <p>; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER</p> <p>; FILE REFERENCE: 210121.561</p> <p>; CURRENT APPLICATION NUMBER: US/09/998,598</p> <p>; CURRENT FILING DATE: 2001-11-16</p> <p>; NUMBER OF SEQ ID NOS: 2606</p> <p>; SOFTWARE: Corixa Invention Disclosure Database</p> <p>; SEQ ID NO: 2240</p> <p>; LENGTH: 413</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Homo sapiens</p> <p>US-09-998-598-2240</p> <p>Query Match 0.5%; Score 20; DB 10; Length 413;</p> <p>Best Local Similarity 100.0%; Pred. No. 2.8;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>									
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935						
Db	364	TTGAGAAAGAGCATGTCGAA	345						
<p>RESULT 9</p> <p>US-09-840-787-54</p> <p>; Sequence 54, Application US/09840787</p> <p>; Patent No. US20020058264A1</p>									

```

;
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Bandman, Olga
; Shah, Purvi
; Au-Young, Janice
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/840,787
; FILING DATE: 23-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/518,865
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1750 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HMCINOT01
; CLONE: 9476
; SEQUENCE DESCRIPTION: SEQ ID NO: 54 :
US-09-840-787-54
;
; Query Match 0.5%; Score 20; DB 10; Length 1750;
; Best Local Similarity 100.0%; Pred. No. 3.2;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1916 TTGAGAAAGAGCATGTGCGAA 1935
;
; Db 1355 TTGAGAAAGAGCATGTGCGAA 1374
;
; RESULT 10
; US-09-782-874-7
; Sequence 7, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:

```

```

;
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-782-874-7
;
; Query Match 0.5%; Score 19; DB 10; Length 34;
; Best Local Similarity 100.0%; Pred. No. 7.3;
; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 33 ACTCATCACTCCCTCAAG 51
;
; Db 16 ACTCATCACTCCCTCAAG 34
;
; RESULT 11
; US-09-563-817-946/C
; Sequence 946, Application US/09563817
; Patent No. US20020095031A1
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael C.
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020095031A1el Human Polynucleotides and the
; FILE REFERENCE: LEX-0021-USA
; CURRENT APPLICATION NUMBER: US/09/563,817
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 60/132,343
; PRIOR FILING DATE: 1999-05-04
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 946
; LENGTH: 267
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(267)
; OTHER INFORMATION: n = A,T,C or G
US-09-563-817-946
;
; Query Match 0.5%; Score 19; DB 10; Length 267;

```

```
Best Local Similarity 100.0%; Pred. No. 9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 TCTTCTTTTGAAGGCTT 428
      |||
Db 22 TCTTCTTTTGAAGGCTT 4

RESULT 12
US-09-864-761-11608/c
; Sequence 11608, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11608
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006204.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.79
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
; US-09-864-761-11608
```

```
Query Match 0.5%; Score 19; DB 10; Length 461;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGGC 647
      |||
Db 56 TCATATGAAATATATGGC 38

RESULT 13
US-09-764-877-2690
; Sequence 2690, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2690
; LENGTH: 4656
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-877-2690
```

```
Query Match 0.5%; Score 19; DB 10; Length 4656;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3567 TATATATGTAATAGGTG 3585
      |||
Db 2822 TATATATGTAATAGGTG 2840
```

```
RESULT 14
US-09-782-874-9/c
; Sequence 9, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassenegeger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RRRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/782,874
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
```

; REFERENCE/DOCKET NUMBER: MFG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-782-874-9

Query Match 0.5%; Score 18; DB 10; Length 35;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3615 GACTACCTTTGTCCTTA 3632
|||||
DB 35 GACTACCTTTGTCCTTA 18

RESULT 15

US-09-923-875-1838/c
; Sequence 1838, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 1838
; LENGTH: 256
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700159613H2
US-09-923-875-1838

Query Match 0.5%; Score 18; DB 10; Length 256;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2120 GAGCGTTTGAATTCATG 2137
|||||
DB 80 GAGCGTTTGAATTCATG 63

RESULT 16

US-09-960-352-1347/c
; Sequence 1347, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1347
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 06-LIB34-055-Q1-E1-B5
US-09-960-352-1347

Query Match 0.5%; Score 18; DB 10; Length 358;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 581 TTTTTCCTTTTACT 598
|||||
DB 263 TTTTTCCTTTTACT 246

RESULT 17

US-09-867-701-4817
; Sequence 4817, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4817
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4817

Query Match 0.5%; Score 18; DB 10; Length 385;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 581 TTTTTCCTTTTACT 598
|||||
DB 114 TTTTTCCTTTTACT 131

RESULT 18

US-09-864-761-11283/c
; Sequence 11283, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/532,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00665

US-09-864-761-11283/c
; Sequence 11283, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/532,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00665

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11283
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004467.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; US-09-864-761-11283

Query Match 0.5%; Score 18; DB 10; Length 474;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2280 TTCCAAATGGAGAACAA 2297
Db 32 TTCCAAATGGAGAACAA 15
|||||
|||||

RESULT 19
US-09-783-590-8123/c
; Sequence 8123, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16-2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8123
; LENGTH: 500
; TYPE: DNA

```



```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (433)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (455)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (461)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (467)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (488)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (473)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (476)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (477)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (479)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (480)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (484)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (486)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (487)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (495)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (497)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-8123
```

```
Query Match          0.5%; Score 18; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 576 AGAAATTTTCTTTT 593
    |||||
```

```
Db 70 AGAAATTTTCTTTT 53
```

RESULT 20

```
; Sequence 176, Application US/09925299
; Patent No. US2002005627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 176
; LENGTH: 1016
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (895)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (928)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (970)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (992)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (1001)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-299-176
```

```
Query Match          0.5%; Score 18; DB 10; Length 1016;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 576 AGAAATTTTCTTTT 593
    |||||
```

```
Db 96 AGAAATTTTCTTTT 79
```

RESULT 21

```
; Sequence 7347, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7347
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1467)
; US-09-815-242-7347
```


RESULT 26

US-09-822-849A-162/c
; Sequence 162, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 162
; LENGTH: 3760
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-162

Query Match 0.5%; Score 18; DB 10; Length 3760;

Best Local Similarity 100.0%; Pred. No. 40;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 AACAGTCCAAAGGAGGA 322

Db 1490 AACAGTCCAAAGGAGGA 1473

RESULT 27

US-10-052-586-453/c
; Sequence 453, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541

; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29

;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/086023
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/086392
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/086486
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087098
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/087208
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/087609
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087759
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087827
;; PRIOR FILING DATE: 1998-06-03
;; PRIOR APPLICATION NUMBER: 60/088025
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088028
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088029
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088722
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088740
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088811
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088825
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088861

;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088863
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089090
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 0.5%; Score 18; DB 12; Length 3781;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3267 GGGCCTTGAGGAAGAGG 3284
|||||
Db 1159 GGGCCTTGAGGAAGAGG 1142

RESULT 28

US-09-879-445-1
; Sequence 1, Application US/09879445
; Patent No. US20020061509A1
; GENERAL INFORMATION:

APPLICANT: Mundy, Gregory R.

TITLE OF INVENTION: SCREENING ASSAY FOR THE IDENTIFICATION
OF AGENTS WHICH INHIBIT CANCER
METASTASIS TO BONE

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,445

FILING DATE: 11-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/915,868

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 37,642

REFERENCE/DOCKET NUMBER: OSTs:002PZ1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4348 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-879-445-1

Query Match 0.5%; Score 18; DB 10; Length 4348;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 576 AGAAATTTTTCCTTTT 593
|||||
DB 4180 AGAAATTTTTCCTTTT 4197

RESULT 29

US-09-764-877-3657
; Sequence 3657, Application US/09764877

; Patent No. US20020147140A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC005

; CURRENT APPLICATION NUMBER: US/09/764,877

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 4031

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3657

; LENGTH: 32192

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-877-3657

Query Match 0.5%; Score 18; DB 10; Length 32192;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TTAAGTCATCTTAGAGA 266
|||||
DB 28605 TTAAGTCATCTTAGAGA 28622

RESULT 30

US-09-927-091-4

; Sequence 4, Application US/09927091

; Patent No. US20020119541A1

; GENERAL INFORMATION:

; APPLICANT: KILLARY, ANN

; APPLICANT: LOTT, STEVE

; APPLICANT: CHANDLER, DAWN

; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1

; FILE REFERENCE: UTSC:65105

; CURRENT APPLICATION NUMBER: US/09/927,091

; CURRENT FILING DATE: 2001-08-09

; PRIOR APPLICATION NUMBER: 60/227,560

; PRIOR FILING DATE: 2000-08-23

; PRIOR APPLICATION NUMBER: 60/225,033

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 49744

; TYPE: DNA

; ORGANISM: Human

US-09-927-091-4

Query Match 0.5%; Score 18; DB 10; Length 49744;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 581 TTTTTCCTTTTATCT 598
|||||
DB 1050 TTTTTCCTTTTATCT 1067

RESULT 31

US-09-835-232-7/c

; Sequence 7, Application US/09835232

; Patent No. US20020098489A1

; GENERAL INFORMATION:

; APPLICANT: Leder, Philip

; APPLICANT: Leader, Benjamin

; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 00383/052002

; CURRENT APPLICATION NUMBER: US/09/835,232

; CURRENT FILING DATE: 2001-04-12

; PRIOR APPLICATION NUMBER: US 60/196,811

; PRIOR FILING DATE: 2000-04-13

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 170834

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (1)...(170834)

; OTHER INFORMATION: n= A,T,C, or G

US-09-835-232-7

Query Match 0.5%; Score 18; DB 10; Length 170834;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 GTTTCAGTGTGGGAAG 540
|||||
DB 36324 GTTTCAGTGTGGGAAG 36307

RESULT 32

US-09-923-876-3754

; Sequence 3754, Application US/09923876

; Patent No. US20020013958A1

; GENERAL INFORMATION:

; APPLICANT: Lalgudi, Raghunath V.

; APPLICANT: Kamigaki, Laura Y. (Ito)

; APPLICANT: Sherman, Bradley K.

; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING

; FILE REFERENCE: PL-0012-1 CON

; CURRENT APPLICATION NUMBER: US/09/923,876

; CURRENT FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: 09/298,329

; PRIOR FILING DATE: 1999-04-21

; PRIOR APPLICATION NUMBER: 60/085,331

; PRIOR FILING DATE: 1998-05-05

; NUMBER OF SEQ ID NOS: 6332

; SOFTWARE: PERL Program

; SEQ ID NO 3754

; LENGTH: 150

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: misc.feature

; OTHER INFORMATION: Incyte ID No. US20020013958A1 700453751H1

; NAME/KEY: unsure

; LOCATION: 115, 118, 121, 129, 138, 147

; OTHER INFORMATION: a, t, c, g, or other

US-09-923-876-3754

Query Match 0.5%; Score 17; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTGATCTG 2480
|||||
DB 73 TCCATGCTTGATCTG 89

```
RESULT 33
US-09-878-574-9084
; Sequence 9084, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 9084
; LENGTH: 213
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701102066H1
US-09-878-574-9084

Query Match          0.5%; Score 17; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2745 TTGTGAATGACAGTTTG 2761
|||||
DB 125 TTGTGAATGACAGTTG 141

RESULT 34
US-09-878-574-7593/c
; Sequence 7593, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 7593
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701100084H2
US-09-878-574-7593

Query Match          0.5%; Score 17; DB 10; Length 258;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2429 TTCATCTGAAGGGGAAA 2445
|||||
DB 66 TTCATCTGAAGGGGAAA 50

RESULT 35
US-09-969-373-41/c
; Sequence 41, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Haug, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
```

```
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US/09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US/09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 41
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Glycine max
US-09-969-373-41

Query Match          0.5%; Score 17; DB 10; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 582 TTTTTCCTTTTATCT 598
|||||
DB 148 TTTTTCCTTTTATCT 132

RESULT 36
US-09-294-093B-1096
; Sequence 1096, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalquidi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 1096
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Inocyte ID No. US20010051335A1 700343770H1
US-09-294-093B-1096

Query Match          0.5%; Score 17; DB 10; Length 274;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2567 CCTCATCCGAATGATG 2583
|||||
DB 128 CCTCATCCGAATGATG 144

RESULT 37
US-09-864-761-31520/c
; Sequence 31520, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31520
; LENGTH: 308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006323.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: NT HIT: US2965.1, EVALUE 1.00e-108
; OTHER INFORMATION: EST HUMAN HIT: BF026728.1, EVALUE 1.00e-108
; OTHER INFORMATION: SWISSPROT HIT: Q15910, EVALUE 6.00e-13
US-09-864-761-31520

Query Match 0.58; Score 17; DB 10; Length 308;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 3712 AGTTTCATCTTTCTCT 3728
DB 228 AGTTTCATCTTTCTCT 212

RESULT 38
US-09-880-107-706/c
; Sequence 706, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Owe

US-09-917-800A-623/c
; Sequence 623, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 623
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI045253
; NAME/KEY: misc.feature
; LOCATION: (1)..(353)
; OTHER INFORMATION: n = a o r c o r g o r t
US-09-917-800A-623
```

Query Match 0.5%; Score 17; DB 10; Length 353;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 AAATTTTCTTTTCTTTT 594
|||||
Db 64 AAATTTTCTTTTCTTTT 48

RESULT 40
US-09-864-761-13956/c
; Sequence 13956, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13956
; LENGTH: 368
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF130350.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
US-09-864-761-13956

Query Match 0.5%; Score 17; DB 10; Length 368;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 TTTATTTTGCATACTT 132
|||||
Db 236 TTTATTTTGCATACTT 220

RESULT 41
US-09-974-300-7289/c
; Sequence 7289, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10095.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7289
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(387)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-7289

Query Match 0.5%; Score 17; DB 10; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 AAATTTTCTTTTCTTTT 594
|||||
Db 268 AAATTTTCTTTTCTTTT 252

RESULT 42
US-09-864-761-30520/c
; Sequence 30520, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 30520
;; LENGTH: 391
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AFL130350.1
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
;; OTHER INFORMATION: NT HIT: ALI63248.2, EVALU0 0.00e+00
;; OTHER INFORMATION: EST_HUMAN HIT: AA001786.1, EVALU0 3.00e-84
;; OTHER INFORMATION: SWISSPROT HIT: P70213, EVALU0 6.00e-21
US-09-864-761-30520

Query Match 0.5%; Score 17; DB 10; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 TTTATTTTGCATAACTT 132
|||||
DB 81 TTTATTTTGCATAACTT 65

RESULT 43
US-09-563-817-878/c
; Sequence 878, Application US/09563817
; Patent No. US20020095031A1
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael C.
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: NO. US20020095031A1 Human Polynucleotides and the
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: LEX-0021-USA
; CURRENT APPLICATION NUMBER: US/09/563,817
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 60/132,343
; PRIOR FILING DATE: 1999-05-04
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 878
; LENGTH: 391
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(391)
; OTHER INFORMATION: n = A,T,C or G

US-09-563-817-878

Query Match 0.5%; Score 17; DB 10; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 CTTTAAGGAACCTCTG 770
|||||
DB 118 CTTTAAGGAACCTCTG 102

RESULT 44

US-09-983-965-401
; Sequence 401, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 401
; LENGTH: 406
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-D12
US-09-983-965-401

Query Match 0.5%; Score 17; DB 10; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2494 TTTAAAGGCTGTAATG 2510
|||||
DB 41 TTTAAAGGCTGTAATG 57

RESULT 45

US-09-867-550-1095
; Sequence 1095, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: NO. US20020082206A1 Polynucleotides from Atherogenic Cells a
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1095
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (336)..(359)

; OTHER INFORMATION: Wherein any n is one of a or t or c or g
US-09-867-550-1095

Query Match 0.5%; Score 17; DB 10; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1508 GATAAAATTTGAATT 1524
Db 201 GATAAAATTTGAATT 217

RESULT 46

US-09-960-352-10698
; Sequence 10698, Application US/09960352
; Patent No. US20020137139A1

GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10698
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 46-LIB3057-002-Q1-K1-D6
US-09-960-352-10698

Query Match 0.5%; Score 17; DB 10; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 580 ATTTTTCCTTTTAT 596
Db 353 ATTTTTCCTTTTAT 369

RESULT 47

US-09-880-107-2705
; Sequence 2705, Application US/09880107
; Patent No. US20020142981A1

GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880.107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2705
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 N81036
US-09-880-107-2705

Query Match 0.5%; Score 17; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1954 AAAGTTAGATGTCTCTG 1970
Db 388 AAAGTTAGATGTCTCTG 404

RESULT 48

US-09-864-761-14986/c
; Sequence 14986, Application US/09864761
; Patent No. US20020048763A1

GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864.761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 14986

LENGTH: 430

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

; OTHER INFORMATION: MAP TO AC006323.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
US-09-864-761-14986

```
Query Match      0.5%; Score 17; DB 10; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3712 AGTTTCATCTTCTTCT 3728
|||||
Db 405 AGTTTCATCTTCTTCT 389

RESULT 49
US-09-864-761-2826
; Sequence 2826, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2826
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AJ229041.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
```

```
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
US-09-864-761-2826

Query Match      0.5%; Score 17; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AAATTTTTCCTTTTA 595
|||||
Db 126 AAATTTTTCCTTTTA 142

RESULT 50
US-09-983-965-2934/c
; Sequence 2934, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2934
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 19-LIB3058-019-Q1-K1-E11
US-09-983-965-2934

Query Match      0.5%; Score 17; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 AAATTTTTCCTTTT 594
|||||
Db 157 AAATTTTTCCTTTT 141

Search completed: November 6, 2002, 03:42:03
Job time : 850 secs
```